

SEQUENCE LISTING



(1) GENERAL INFORMATION:

- (i) APPLICANT: FLECKENSTEIN, Bernhard
ENSSER, Armin
- (ii) TITLE OF INVENTION: HUMAN SEMAPHORIN L (H-SEMAL) AND
CORRESPONDING SEMAPHORINS IN OTHER SPECIES
- (iii) NUMBER OF SEQUENCES: 44
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Frommer Lawrence & Haug LLP
 - (B) STREET: 745 Fifth Avenue
 - (C) CITY: New York
 - (D) STATE: New York
 - (E) COUNTRY: USA
 - (F) ZIP: 10151
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US NYA
 - (B) FILING DATE: 09-JUL-1998
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Lawrence, William F.
 - (B) REGISTRATION NUMBER: 28,029
 - (C) REFERENCE/DOCKET NUMBER: 514429-3647
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 212-588-0800
 - (B) TELEFAX: 212-588-0500

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 2636 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CGGGGCCACG GGATGACGCC TCCTCCGCCC GGACGTGCCG CCCCAGCGC ACCGCGCGCC	60
CGCGTCCCTG GCGCGCCGGC TCGGTTGGGG CTTCCGCTGC GGCTGCGGCT GCTGCTGCTG	120
CTCTGGGCGG CCGCCGCCTC CGCCCAGGGC CACCTAAGGA GCGGACCCCG CATCTTCGCC	180
GTCTGGAAAG GCCATGTAGG GCAGGACCGG GTGGACTTTG GCCAGACTGA GCCGCACACG	240
GTGCTTTTCC ACGAGCCAGG CAGCTCCTCT GTGTGGGTGG GAGGACGTGG CAAGGTCTAC	300
CTCTTTGACT TCCCCGAGGG CAAGAACGCA TCTGTGCGCA CCGTGAATAT CGGCTCCACA	360
AAGGGGTCCT GTCTGGATAA GCGGGACTGC GAGAACTACA TCACTCTCCT GGAGAGGCGG	420
AGTGAGGGGC TGCTGGCCTG TGGCACCAAC GCGCGCACC CCAGCTGCTG GAACCTGGTG	480
AATGGCACTG TGGTGCCACT TGGCGAGATG AGAGGCTACG CCCCCTTCAG CCCGGACGAG	540
AACTCCCTGG TTCTGTTTGA AGGGGACGAG GTGTATTCCA CCATCCGGAA GCAGGAATAC	600
AATGGGAAGA TCCCTCGGTT CCGCCGCATC CGGGGCGAGA GTGAGCTGTA CACCAGTGAT	660
ACTGTCATGC AGAACCACA GTTCATCAA GCCACCATCG TGCACCAAGA CCAGGCTTAC	720
GATGACAAGA TCTACTACTT CTTCCGAGAG GACAATCCTG ACAAGAATCC TGAGGCTCCT	780
CTCAATGTGT CCCGTGTGGC CCAGTTGTGC AGGGGGGACC AGGGTGGGGA AAGTTCACTG	840
TCAGTCTCCA AGTGGAACAC TTTTCTGAAA GCCATGCTGG TATGCAGTGA TGCTGCCACC	900
AACAAGAACT TCAACAGGCT GCAAGACGTC TTCCTGCTCC CTGACCCAG CGGCCAGTGG	960
AGGGACACCA GGGTCTATGG TGTTTTCTCC AACCCTGGA ACTACTCAGC CGTCTGTGTG	1020
TATTCCTCG GTGACATTGA CAAGGTCTTC CGTACCTCCT CACTCAAGGG CTACCACTCA	1080
AGCCTTCCCA ACCCGCGGCC TGGCAAGTGC CTCCAGACC AGCAGCCGAT ACCCACAGAG	1140
ACCTTCCAGG TGGCTGACCG TCACCCAGAG GTGGCGCAGA GGGTGGAGCC CATGGGGCCT	1200
CTGAAGACGC CATTGTTCCA CTCTAAATAC CACTACCAGA AAGTGGCCGT TCACCGCATG	1260
CAAGCCAGCC ACGGGGAGAC CTTTCATGTG CTTTACCTAA CTACAGACAG GGGCACTATC	1320
CACAAGGTGG TGGAACCGGG GGAGCAGGAG CACAGCTTCG CCTTCAACAT CATGGAGATC	1380
CAGCCCTTCC GCCGCGCGGC TGCCATCCAG ACCATGTCGC TGGATGCTGA GCGGAGGAAG	1440
CTGTATGTGA GCTCCAGTG GGAGGTGAGC CAGGTGCCCC TGGACCTGTG TGAGGTCTAT	1500
GGCGGGGGCT GCCACGGTTG CCTCATGTCC CGAGACCCCT ACTGCGGCTG GGACCAGGGC	1560
CGCTGCATCT CCATCTACAG CTCCGAACGG TCAGTGCTGC AATCCATTAA TCCAGCCGAG	1620
CCACACAAGG AGTGTCCCAA CCCCAAACCA GACAAGGCC CACTGCAGAA GGTTTCCCTG	1680
GCCCCAACT CTCGCTACTA CCTGAGCTGC CCCATGGAAT CCCGCCACGC CACCTACTCA	1740

TGGCGCCACA AGGAGAACGT GGAGCAGAGC TGCGAACCTG GTCACCAGAG CCCCAACTGC	1800
ATCCTGTTCA TCGAGAACCT CACGGCGCAG CAGTACGGCC ACTACTTCTG CGAGGCCCCAG	1860
GAGGGCTCCT ACTTCCGCGA GGCTCAGCAC TGGCAGCTGC TGCCCGAGGA CGGCATCATG	1920
GCCGAGCACC TGCTGGGTCA TGCCTGTGCC CTGGCTGCCT CCCTCTGGCT GGGGGTGCTG	1980
CCCACACTCA CTCTTGGCTT GCTGGTCCAC TAGGGCCTCC CGAGGCTGGG CATGCCTCAG	2040
GCTTCTGCAG CCCAGGGCAC TAGAACGTCT CACACTCAGA GCCGGCTGGC CCGGGAGCTC	2100
CTTGCCCTGCC ACTTCTTCCA GGGGACAGAA TAACCCAGTG GAGGATGCCA GGCCTGGAGA	2160
CGTCCAGCCG CAGGCGGCTG CTGGGCCCCA GGTGGCGCAC GGATGGTGAG GGGCTGAGAA	2220
TGAGGGCACC GACTGTGAAG CTGGGGCATC GATGACCCAA GACTTTATCT TCTGGAAAAT	2280
ATTTTTCAGA CTCCTCAAAC TTGACTAAAT GCAGCGATGC TCCCAGCCCA AGAGCCCATG	2340
GGTCGGGGAG TGGGTTTGGA TAGGAGAGCT GGGACTCCAT CTCGACCCTG GGGCTGAGGC	2400
CTGAGTCCTT CTGGACTCTT GGTACCCACA TTGCCTCCTT CCCCTCCCTC TCTCATGGCT	2460
GGGTGGCTGG TGTTCCTGAA GACCCAGGGC TACCCTCTGT CCAGCCCTGT CCTCTGCAGC	2520
TCCCTCTCTG GTCCTGGGTC CCACAGGACA GCCGCCTTGC ATGTTTATTG AAGGATGTTT	2580
GCTTTCCGGA CGGAAGGACG GAAAAAGCTC TGAAAAAAAA AAAAAAAAAA AAAAAA	2636

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1195 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CGGGGCTGCG GGATGACGCC TCCTCTCCC GGACGTGCCG CCCCCAGCGC ACCGCGCGCC	60
CGCGTCCTCA GCCTGCCGGC TCGGTTCTGG CTCCCGCTGC GGCTGCGGCT TCTGCTGGTG	120
TTCTGGGTGG CCGCCGCCTC CGCCCAAGGC CACTCGAGGA GCGGACCCCG CATCTCCGCC	180
GTCTGGAAAG GGCAGGACCA TGTGGACTTT AGCCAGCCTG AGCCACACAC CGTGCTTTTC	240
CATGAGCCGG GCAGCTTCTC TGTCTGGGTG GGTGGACGTG GCAAGGTCTA CCACTTCAAC	300
TTCCCCGAGG GCAAGAATGC CTCTGTGCGC ACGGTGAACA TCGGCTCCAC AAAGGGGTCC	360

TGTCAGGACA AACAGGACTG TGGGAATTAC ATCACTCTTC TAGAAAGGCG GGGTAATGGG	420
CTGCTGGTCT GTGGCACCAA TGCCCGGAAG CCCAGCTGCT GGAAGTTGGT GAATGACAGT	480
GTGGTGATGT CACTTGGTGA GATGAAAGGC TATGCCCCCT TCAGCCCGGA TGAGAACTCC	540
CTGGTTCTGT TTGAAGGAGA TGAAGTGTAC TCTACCATCC GGAAGCAGGA ATACAACGGG	600
AAGATCCCTC GGTTTCGACG CATTTCGGGGC GAGAGTGAAC TGTACACAAG TGATACAGTC	660
ATGCAGAACC CACAGTTCAT CAAGGCCACC ATTGTGCACC AAGACCAAGC CTATGATGAT	720
AAGATCTACT ACTTCTTCCG AGAAGACAAC CCTGACAAGA ACCCCGAGGC TCCTCTCAAT	780
GTGTCCCGAG TAGCCAGTT GTGCAGGGGG GACCAGGGTG GTGAGAGTTC GTTGTCTGTC	840
TCCAAGTGGA ACACCTTCCT GAAAGCCATG TTGGTCTGCA GCGATGCAGC CACCAACAGG	900
AACTTCAATC GGCTGCAAGA TGTCTTCCTG CTCCCTGACC CCAGTGGCCA GTGGAGAGAT	960
ACCAGGGTCT ATGGCGTTTT CTCCAACCCC TGGAATACT CAGCTGTCTG CGTGTATTCG	1020
CTTGGTGACA TTGACAGAGT CTTCCGTACC TCATCGCTCA AAGGCTACCA CATGGGCCTT	1080
TCCAACCCTC GACCTGGCAT GTGCCTCCCA AAAAAGCAGC CCATACCCAC AGAAACCTTC	1140
CAGGTAGCTG ATAGTCACCC AGAGGTGGCT CAGAGGGTGG AACCTATGGG GCCCC	1195

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 666 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: n/a
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: amino acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met	Thr	Pro	Pro	Pro	Pro	Gly	Arg	Ala	Ala	Pro	Ser	Ala	Pro	Arg	Ala
1				5					10					15	

Arg	Val	Pro	Gly	Pro	Pro	Ala	Arg	Leu	Gly	Leu	Pro	Leu	Arg	Leu	Arg
			20					25					30		

Leu	Leu	Leu	Leu	Leu	Trp	Ala	Ala	Ala	Ala	Ser	Ala	Gln	Gly	His	Leu
			35				40						45		

Arg	Ser	Gly	Pro	Arg	Ile	Phe	Ala	Val	Trp	Lys	Gly	His	Val	Gly	Gln
	50						55					60			

Asp	Arg	Val	Asp	Phe	Gly	Gln	Thr	Glu	Pro	His	Thr	Val	Leu	Phe	His
-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

65				70				75				80			
Glu	Pro	Gly	Ser	Ser	Ser	Val	Trp	Val	Gly	Gly	Arg	Gly	Lys	Val	Tyr
				85					90					95	
Leu	Phe	Asp	Phe	Pro	Glu	Gly	Lys	Asn	Ala	Ser	Val	Arg	Thr	Val	Asn
			100					105					110		
Ile	Gly	Ser	Thr	Lys	Gly	Ser	Cys	Leu	Asp	Lys	Arg	Asp	Cys	Glu	Asn
		115					120					125			
Tyr	Ile	Thr	Leu	Leu	Glu	Arg	Arg	Ser	Glu	Gly	Leu	Leu	Ala	Cys	Gly
		130				135					140				
Thr	Asn	Ala	Arg	His	Pro	Ser	Cys	Trp	Asn	Leu	Val	Asn	Gly	Thr	Val
145					150					155					160
Val	Pro	Leu	Gly	Glu	Met	Arg	Gly	Tyr	Ala	Pro	Phe	Ser	Pro	Asp	Glu
				165					170					175	
Asn	Ser	Leu	Val	Leu	Phe	Glu	Gly	Asp	Glu	Val	Tyr	Ser	Thr	Ile	Arg
			180					185					190		
Lys	Gln	Glu	Tyr	Asn	Gly	Lys	Ile	Pro	Arg	Phe	Arg	Arg	Ile	Arg	Gly
		195					200					205			
Glu	Ser	Glu	Leu	Tyr	Thr	Ser	Asp	Thr	Val	Met	Gln	Asn	Pro	Gln	Phe
		210				215					220				
Ile	Lys	Ala	Thr	Ile	Val	His	Gln	Asp	Gln	Ala	Tyr	Asp	Asp	Lys	Ile
225					230					235					240
Tyr	Tyr	Phe	Phe	Arg	Glu	Asp	Asn	Pro	Asp	Lys	Asn	Pro	Glu	Ala	Pro
				245					250					255	
Leu	Asn	Val	Ser	Arg	Val	Ala	Gln	Leu	Cys	Arg	Gly	Asp	Gln	Gly	Gly
			260					265					270		
Glu	Ser	Ser	Leu	Ser	Val	Ser	Lys	Trp	Asn	Thr	Phe	Leu	Lys	Ala	Met
		275					280					285			
Leu	Val	Cys	Ser	Asp	Ala	Ala	Thr	Asn	Lys	Asn	Phe	Asn	Arg	Leu	Gln
						295					300				
Asp	Val	Phe	Leu	Leu	Pro	Asp	Pro	Ser	Gly	Gln	Trp	Arg	Asp	Thr	Arg
305					310					315					320
Val	Tyr	Gly	Val	Phe	Ser	Asn	Pro	Trp	Asn	Tyr	Ser	Ala	Val	Cys	Val
				325					330					335	
Tyr	Ser	Leu	Gly	Asp	Ile	Asp	Lys	Val	Phe	Arg	Thr	Ser	Ser	Leu	Lys
			340					345					350		
Gly	Tyr	His	Ser	Ser	Leu	Pro	Asn	Pro	Arg	Pro	Gly	Lys	Cys	Leu	Pro
		355					360					365			
Asp	Gln	Gln	Pro	Ile	Pro	Thr	Glu	Thr	Phe	Gln	Val	Ala	Asp	Arg	His

370	375	380
Pro Glu Val Ala Gln Arg Val Glu Pro Met Gly Pro Leu Lys Thr Pro		
385	390	395 400
Leu Phe His Ser Lys Tyr His Tyr Gln Lys Val Ala Val His Arg Met		
	405	410 415
Gln Ala Ser His Gly Glu Thr Phe His Val Leu Tyr Leu Thr Thr Asp		
	420	425 430
Arg Gly Thr Ile His Lys Val Val Glu Pro Gly Glu Gln Glu His Ser		
	435	440 445
Phe Ala Phe Asn Ile Met Glu Ile Gln Pro Phe Arg Arg Ala Ala Ala		
	450	455 460
Ile Gln Thr Met Ser Leu Asp Ala Glu Arg Arg Lys Leu Tyr Val Ser		
465	470	475 480
Ser Gln Trp Glu Val Ser Gln Val Pro Leu Asp Leu Cys Glu Val Tyr		
	485	490 495
Gly Gly Gly Cys His Gly Cys Leu Met Ser Arg Asp Pro Tyr Cys Gly		
	500	505 510
Trp Asp Gln Gly Arg Cys Ile Ser Ile Tyr Ser Ser Glu Arg Ser Val		
	515	520 525
Leu Gln Ser Ile Asn Pro Ala Glu Pro His Lys Glu Cys Pro Asn Pro		
	530	535 540
Lys Pro Asp Lys Ala Pro Leu Gln Lys Val Ser Leu Ala Pro Asn Ser		
545	550	555 560
Arg Tyr Tyr Leu Ser Cys Pro Met Glu Ser Arg His Ala Thr Tyr Ser		
	565	570 575
Trp Arg His Lys Glu Asn Val Glu Gln Ser Cys Glu Pro Gly His Gln		
	580	585 590
Ser Pro Asn Cys Ile Leu Phe Ile Glu Asn Leu Thr Ala Gln Gln Tyr		
	595	600 605
Gly His Tyr Phe Cys Glu Ala Gln Glu Gly Ser Tyr Phe Arg Glu Ala		
	610	615 620
Gln His Trp Gln Leu Leu Pro Glu Asp Gly Ile Met Ala Glu His Leu		
625	630	635 640
Leu Gly His Ala Cys Ala Leu Ala Ala Ser Leu Trp Leu Gly Val Leu		
	645	650 655
Pro Thr Leu Thr Leu Gly Leu Leu Val His		
	660	665

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 394 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: n/a

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: amino acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Thr Pro Pro Pro Gly Arg Ala Ala Pro Ser Ala Pro Arg Ala
1 5 10 15
Arg Val Leu Ser Leu Pro Ala Arg Phe Gly Leu Pro Leu Arg Leu Arg
20 25 30
Leu Leu Leu Val Phe Trp Val Ala Ala Ala Ser Ala Gln Gly His Ser
35 40 45
Arg Ser Gly Pro Arg Ile Ser Ala Val Trp Lys Gly Gln Asp His Val
50 55 60
Asp Phe Ser Gln Pro Glu Pro His Thr Val Leu Phe His Glu Pro Gly
65 70 75 80
Ser Phe Ser Val Trp Val Gly Gly Arg Gly Lys Val Tyr His Phe Asn
85 90 95
Phe Pro Glu Gly Lys Asn Ala Ser Val Arg Thr Val Asn Ile Gly Ser
100 105 110
Thr Lys Gly Ser Cys Gln Asp Lys Gln Asp Cys Gly Asn Tyr Ile Thr
115 120 125
Leu Leu Glu Arg Arg Gly Asn Gly Leu Leu Val Cys Gly Thr Asn Ala
130 135 140
Arg Lys Pro Ser Cys Trp Asn Leu Val Asn Asp Ser Val Val Met Ser
145 150 155 160
Leu Gly Glu Met Lys Gly Tyr Ala Pro Phe Ser Pro Asp Glu Asn Ser
165 170 175
Leu Val Leu Phe Glu Gly Asp Glu Val Tyr Ser Thr Ile Arg Lys Gln
180 185 190
Glu Tyr Asn Gly Lys Ile Pro Arg Phe Arg Arg Ile Arg Gly Glu Ser
195 200 205
Glu Leu Tyr Thr Ser Asp Thr Val Met Gln Asn Pro Gln Phe Ile Lys
210 215 220
Ala Thr Ile Val His Gln Asp Gln Ala Tyr Asp Asp Lys Ile Tyr Tyr

225		230		235		240
Phe Phe Arg Glu Asp Asn Pro Asp Lys Asn Pro Glu Ala Pro Leu Asn						
	245			250		255
Val Ser Arg Val Ala Gln Leu Cys Arg Gly Asp Gln Gly Gly Glu Ser						
	260			265		270
Ser Leu Ser Val Ser Lys Trp Asn Thr Phe Leu Lys Ala Met Leu Val						
	275			280		285
Cys Ser Asp Ala Ala Thr Asn Arg Asn Phe Asn Arg Leu Gln Asp Val						
	290			295		300
Phe Leu Leu Pro Asp Pro Ser Gly Gln Trp Arg Asp Thr Arg Val Tyr						
305		310		315		320
Gly Val Phe Ser Asn Pro Trp Asn Tyr Ser Ala Val Cys Val Tyr Ser						
	325			330		335
Leu Gly Asp Ile Asp Arg Val Phe Arg Thr Ser Ser Leu Lys Gly Tyr						
	340			345		350
His Met Gly Leu Ser Asn Pro Arg Pro Gly Met Cys Leu Pro Lys Lys						
	355			360		365
Gln Pro Ile Pro Thr Glu Thr Phe Gln Val Ala Asp Ser His Pro Glu						
	370			375		380
Val Ala Gln Arg Val Glu Pro Met Gly Pro						
385		390				

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ACTCACTATA GGGCTCGAGC GGC

23

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AGCCGCACAC GGTGCTTTTC

20

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

GCACAGATGC GTTCTTGCCC

20

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ACCATAGACC CTGGTGTCCC

20

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GCAGTGATGC TGCCACCAAC

20

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CCAGACCATG TCGCTGGATG

20

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

ACATGAGGCA ACCGTGGCAG

20

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CCATCCTAAT ACGACTCACT ATAGGGC

27

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

AGGTAGACCT TGCCACGTCC

20

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GAACTTCAAC AGGCTGCAAG ACG

23

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

ATGCTGAGCG GAGGAAGCTG

20

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CCGCCATACA CCTCACACAG

20

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CTGGAAGCTT TCTGTGGGTA TCGGCTGC

28

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TTTGGATCCC TGGTTCTGTT TGAAG

25

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TTCTAGAATT CAGCGGCCGC TTTTTTTTTT TTTTTTTTTT TTTTTTTTTT

50

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GGGGAAGTT CACTGTCAGT CTCCAAG

27

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GGGAATACAC ACAGACGGCT GAGTAG

26

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

AGCAAGTTCA GCCTGGTTAA GT

22

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TTATGAGTAT TTCTTCCAGG G

21

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CCATTAATCC AGCCGAGCCA CACAAG

26

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CATCTACAGC TCCGAACGGT CAGTG

25

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CAGCGGAAGC CCAACCGAG

20

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GGGATGACGC CTCCTCCGCC CGG

23

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 31 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

AAGCTTCACG TGGACCAGCA AGCCAAGAGT G

31

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

AAGCTTTTTTC CGTCCTCCG TCCGG

25

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

ATGGTGAGCA AGGGCGAGGA GCTG

24

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CTTGTCAGC TCGTCCATGC CGAG

24

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GGGTGGTGAG AGTTCGTTGT CTGTC

25

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 25 base pairs
- (B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GAGCGATGAG GTACGGAAGA CTCTG

25

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5856 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

AGCGCCCAAT ACGCAAACCG CCTCTCCCCG CGCGTTGGCC GATTCATTAA TGCAGCTGGC	60
ACGACAGGTT TCCCGACTGG AAAGCGGGCA GTGAGCGCAA CGCAATTAAT GTGAGTTAGC	120
TCACTCATTA GGCACCCCAG GCTTTACACT TTATGCTTCC GGCTCGTATG TTGTGTGGAA	180
TTGTGAGCGG ATAACAATTT CACACAGGAA ACAGCTATGA CCATGATTAC GCCAAGCTTC	240
ACGTGGACCA GCAAGCCAAG AGTGAGTGTG GGCAGCACCC CCAGCCAGAG GGAGGCAGCC	300
AGGGCACAGG CATGACCCAG CAGGTGCTCG GCCATGATGC CGTCCTCGGG CAGCAGCTGC	360
CAGTGCTGAG CCTCGCGGAA GTAGGAGCCC TCCTGGGCCT CGCAGAAGTA GTGGCCGTAC	420
TGCTGCGCCG TGAGGTTCTC GATGAACAGG ATGCAGTTGG GGCTCTGGTG ACCAGGTTCTG	480
CAGCTCTGCT CCACGTTCTC CTTGTGGCGC CATGAGTAGG TGGCGTGGCG GGATTCCATG	540
GGGACAGCTCA GGTAGTAGCG AGAGTTTGGG GCCAGGGAAA CCTTCTGCAG TGGGGCCTTG	600
TCTGGTTTGG GGTGGGACA CTCCTTGTGT GGCTCGGCTG GATTAATGGA TTGCAGCACT	660
GACCGTTCCG AGCTGTAGAT GGAGATGCAG CGGCCCTGGT CCCAGCCGCA GTAGGGGTCT	720
CGGGACATGA GGCAACCGTG GCAGCCCCCG CCATAGACCT CACACAGGTC CAGGGGCACC	780
TGGCTCACCT CCCACTGGGA GCTCACATAC AGCTTCCTCC GCTCAGCATC CAGCGACATG	840
GTCTGGATGG CAGCCGCGCG GCGGAAGGGC TGGATCTCCA TGATGTTGAA GGCGAAGCTG	900

TCGGTCGCCG GGCGCGGTAT TCTCAGAATG ACTTG GTTGA GTACTCACCA GTCACAGAAA	4380
AGCATCTTAC GGATGGCATG ACAGTAAGAG AATTATGCAG TGCTGCCATA ACCATGAGTG	4440
ATAACACTGC GGCCAACTTA CTTCTGACAA CGATCGGAGG ACCGAAGGAG CTAACCGCTT	4500
TTTTCACAA CATGGGGGAT CATGTA ACTC GCCTTGATCG TTGGGAACCG GAGCTGAATG	4560
AAGCCATACC AAACGACGAG AGTGACACCA CGATGCCTGT AGCAATGCCA ACAACGTTGC	4620
GCAAACTATT AACTGGCGAA CTACTTACTC TAGCTTCCCG GCAACAATTA ATAGACTGGA	4680
TGGAGGCGGA TAAAGTTGCA GGACCACTTC TGCCTCGGC CCTTCCGGCT GGCTGGTTTA	4740
TTGCTGATAA ATCTGGAGCC GGTGAGCGTG GGTCTCGCGG TATCATTGCA GCACTGGGGC	4800
CAGATGGTAA GCCCTCCCGT ATCGTAGTTA TCTACACGAC GGGGAGTCAG GCAACTATGG	4860
ATGAACGAAA TAGACAGATC GCTGAGATAG GTGCCTCACT GATTAAGCAT TGGTAACTGT	4920
CAGACCAAGT TTAATCATAT AACTTTTAGA TTGATTTAAA ACTTCATTTT TAATTTAAAA	4980
GGATCTAGGT GAAGATCCTT TTTGATAATC TCATGACCAA AATCCCTTAA CGTGAGTTTT	5040
CGTTCCACTG AGCGTCAGAC CCCGTAGAAA AGATCAAAGG ATCTTCTTGA GATCCTTTTT	5100
TTCTGCGCGT AATCTGCTGC TTGCAAACAA AAAAACCACC GCTACCAGCG GTGGTTTGTT	5160
TGCCGGATCA AGAGCTACCA ACTCTTTTTTC CGAAGGTAAC TGGCTTCAGC AGAGCGCAGA	5220
TACCAAATAC TGTCTTCTA GTGTAGCCGT AGTTAGGCCA CCACTTCAAG AACTCTGTAG	5280
CACCGCCTAC ATACCTCGCT CTGCTAATCC TGTTACCAGT GGCTGCTGCC AGTGGCGATA	5340
AGTCGTGTCT TACCGGGTTG GACTCAAGAC GATAGTTACC GGATAAGGCG CAGCGGTCGG	5400
GCTGAACGGG GGGTTCGTGC ACACAGCCCA GCTTGGAGCG AACGACCTAC ACCGA ACTGA	5460
GATACCTACA GCGTGAGCAT TGAGAAAGCG CCACGCTTCC CGAAGGGAGA AAGGCGGACA	5520
GGTATCCGGT AAGCGGCAGG GTCGGAACAG GAGAGCGCAC GAGGGAGCTT CCAGGGGGAA	5580
ACGCCTGGTA TCTTTATAGT CCTGTGCGGT TTCGCCACCT CTGACTTGAG CGTCGATTTT	5640
TGTGATGCTC GTCAGGGGGG CGGAGCCTAT GGAAAAACGC CAGCAACGCG GCCTTTTTAC	5700
GGTTCCTGGC CTTTGTCTGG CCTTTGCTC ACATGTTCTT TCCTGCGTTA TCCCCTGATT	5760
CTGTGGATAA CCGTATTACC GCCTTTGAGT GAGCTGATAC CGCTCGCCGC AGCCGAACGA	5820
CCGAGCGCAG CGAGTCAGTG AGCGAGGAAG CGGAAG	5856

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7475 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

GACGGATCGG GAGATCTCCC GATCCCCTAT GGTGCGACTCT CAGTACAATC TGCTCTGATG	60
CCGCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTGTGTT GGAGGTCGCT GAGTAGTGCG	120
CGAGCAAAAT TTAAGCTACA ACAAGGCAAG GCTTGACCGA CAATTGCATG AAGAATCTGC	180
TTAGGGTTAG GCGTTTTGCG CTGCTTCGCG ATGTACGGGC CAGATATACG CGTTGACATT	240
GATTATTGAC TAGTTATTAA TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA	300
TGGAGTTCCG CGTTACATAA CTTACGGTAA ATGGCCCGCC TGGCTGACCG CCCAACGACC	360
CCCGCCCATT GACGTCAATA ATGACGTATG TTCCCATAGT AACGCCAATA GGGACTTTCC	420
ATTGACGTCA ATGGGTGGAC TATTTACGGT AAAGTGGCCA CTTGGCAGTA CATCAAGTGT	480
ATCATATGCC AAGTACGCC CCTATTGACG TCAATGACGG TAAATGGCCC GCCTGGCATT	540
ATGCCCAGTA CATGACCTTA TGGGACTTTC CTAAGTGGCA GTACATCTAC GTATTAGTCA	600
TCGCTATTAC CATGGTGATG CGGTTTTGGC AGTACATCAA TGGGCGTGGA TAGCGGTTTG	660
ACTCACGGGG ATTTCCAAGT CTCCACCCCA TTGACGTCAA TGGGAGTTTG TTTTGGCACC	720
AAAATCAACG GGAAGTTTCCA AAATGTCGTA ACAACTCCGC CCCATTGACG CAAATGGGCG	780
GTAGGCGTGT ACGGTGGGAG GTCTATATAA GCAGAGCTCT CTGGCTAACT AGAGAACCCA	840
CTGCTTACTG GCTTATCGAA ATTAATACGA CTCACTATAG GGAGACCCAA GCTGGCTAGC	900
GTTTAAACGG GCCCTCTAGA CTCGAGCGGC CGCCACTGTG CTGGATATCT GCAGAATTCTG	960
GCTTGGGATG ACGCCTCCTC CGCCCGGACG TGCCGCCCCC AGCGCACCGC GCGCCCGCGT	1020
CCCTGGCCCCG CCGGCTCGGT TGGGGCTTCC GCTGCGGCTG CGGCTGCTGC TGCTGCTCTG	1080
GGCGGCCGCC GCCTCCGCCC AGGGCCACCT AAGGAGCGGA CCCCAGATCT TCGCCGTCTG	1140
GAAAGGCCAT GTAGGGCAGG ACCGGGTGGA CTTTGGCCAG ACTGAGCCGC ACACGGTGCT	1200
TTTCCACGAG CCAGGCAGCT CCTCTGTGTG GGTGGGAGGA CGTGGCAAGG TCTACCTCTT	1260
TGACTTCCCC GAGGGCAAGA ACGCATCTGT GCGCACGGTG AATATCGGCT CCACAAAGGG	1320
GTCCTGTCTG GATAAGCGGG ACTGCGAGAA CTACATCACT CTCCTGGAGA GCGGGAGTGA	1380
GGGGCTGCTG GCCTGTGGCA CCAACGCCCC GCACCCAGC TGCTGGAACC TGGTGAATGG	1440

TCCTTGACCC TGGAAGGTGC CACTCCCCT GTCCTTTCT AATAAAATGA GGAAATTGCA	3180
TCGCATTGTC TGAGTAGGTG TCATTCTATT CTGGGGGGTG GGGTGGGGCA GGACAGCAAG	3240
GGGGAGGATT GGGAAGACAA TAGCAGGCAT GCTGGGGATG CGGTGGGCTC TATGGCTTCT	3300
GAGGCGGAAA GAACCAGCTG GGGCTCTAGG GGGTATCCCC ACGCGCCCTG TAGCGGCGCA	3360
TTAAGCGCGG CGGGTGTGGT GGTACGCGC AGCGTGACCG CTACACTTGC CAGCGCCCTA	3420
GCGCCCGCTC CTTTCGCTTT CTTCCTTCC TTTCTCGCCA CGTTCGCCGG CTTTCCCCGT	3480
CAAGCTCTAA ATCGGGGCAT CCCTTTAGGG TTCCGATTTA GTGCTTTACG GCACCTCGAC	3540
CCCAAAAAAC TTGATTAGGG TGATGGTTCA CGTAGTGGGC CATCGCCCTG ATAGACGGTT	3600
TTTCGCCCTT TGACGTTGGA GTCCACGTTT TTTAATAGTG GACTCTTGTT CCAAACCTGGA	3660
ACAACACTCA ACCCTATCTC GGTCTATTCT TTTGATTTAT AAGGGATTTT GGGGATTTCG	3720
GCCTATTGGT TAAAAAATGA GCTGATTTAA CAAAAATTTA ACGCGAATTA ATTCTGTGGA	3780
ATGTGTGTCA GTTAGGGTGT GGAAAGTCCC CAGGCTCCCC AGGCAGGCAG AAGTATGCAA	3840
AGCATGCATC TCAATTAGTC AGCAACCAGG TGTGGAAAGT CCCCAGGCTC CCCAGCAGGC	3900
AGAAGTATGC AAAGCATGCA TCTCAATTAG TCAGCAACCA TAGTCCCGCC CCTAACTCCG	3960
CCCATCCGCG CCCTAACTCC GCCCAGTTCC GCCCATTCTC CGCCCCATGG CTGACTAATT	4020
TTTTTTATTT ATGCAGAGGC CGAGGCCGCC TCTGCCTCTG AGCTATTCCA GAAGTAGTGA	4080
GGAGGCTTTT TTGGAGGCCT AGGCTTTTGC AAAAAGCTCC CGGGAGCTTG TATATCCATT	4140
TTCGGATCTG ATCAAGAGAC AGGATGAGGA TCGTTTCGCA TGATTGAACA AGATGGATTG	4200
CACGCAGGTT CTCCGGCCGC TTGGGTGGAG AGGCTATTCG GCTATGACTG GGCACAACAG	4260
ACAATCGGCT GCTCTGATGC CGCCGTGTTC CGGCTGTCAG CGCAGGGGCG CCCGGTTCTT	4320
TTTGTCAAGA CCGACCTGTC CGGTGCCCTG AATGAACTGC AGGACGAGGC AGCGCGGCTA	4380
TCGTGGCTGG CCACGACGGG CGTTCCTTGC GCAGCTGTGC TCGACGTTGT CACTGAAGCG	4440
GGAAGGGACT GGCTGCTATT GGGCGAAGTG CCGGGGCAGG ATCTCCTGTC ATCTCACCTT	4500
GCTCCTGCCG AGAAAGTATC CATCATGGCT GATGCAATGC GCGGGCTGCA TACGCTTGAT	4560
CCGGCTACCT GCCCATTGCA CCACCAAGCG AAACATCGCA TCGAGCGAGC ACGTACTCGG	4620
ATGGAAGCCG GTCTTGTCGA TCAGGATGAT CTGGACGAAG AGCATCAGGG GCTCGCGCCA	4680
GCCGAACTGT TCGCCAGGCT CAAGGCGCGC ATGCCCCGACG GCGAGGATCT CGTCGTGACC	4740
CATGGCGATG CCTGCTTGCC GAATATCATG GTGGAAAATG GCCGCTTTTC TGGATTTCATC	4800
GACTGTGGCC GGCTGGGTGT GGCGGACCGC TATCAGGACA TAGCGTTGGC TACCCGTGAT	4860

ATTGCTGAAG AGCTTGGCGG CGAATGGGCT GACCGCTTCC TCGTGCTTTA CGGTATCGCC	4920
GCTCCCGATT CGCAGCGCAT CGCCTTCTAT CGCCTTCTTG ACGAGTTCTT CTGAGCGGGA	4980
CTCTGGGGTT CGAAATGACC GACCAAGCGA CGCCCAACCT GCCATCACGA GATTTTCGATT	5040
CCACCGCCGC CTTCTATGAA AGGTTGGGCT TCGGAATCGT TTTCCGGGAC GCCGGCTGGA	5100
TGATCCTCCA GCGCGGGGAT CTCATGCTGG AGTTCTTCGC CCACCCCAAC TTGTTTATTG	5160
CAGCTTATAA TGGTTACAAA TAAAGCAATA GCATCACAAA TTTCACAAAT AAAGCATTTT	5220
TTTCACTGCA TTCTAGTTGT GGTTCGTCCA AACTCATCAA TGTATCTTAT CATGTCTGTA	5280
TACCGTCGAC CTCTAGCTAG AGCTTGGCGT AATCATGGTC ATAGCTGTTT CCTGTGTGAA	5340
ATTGTTATCC GCTCACAATT CCACACAACA TACGAGCCGG AAGCATAAAG TGTAAAGCCT	5400
GGGGTGCCTA ATGAGTGAGC TAACTCACAT TAATTGCGTT GCGCTCACTG CCCGCTTTC	5460
AGTCGGGAAA CCTGTCTGTC CAGCTGCATT AATGAATCGG CCAACGCGCG GGGAGAGGCG	5520
GTTTGCGTAT TGGGCGCTCT TCCGCTTCTT CGCTCACTGA CTCGCTGCGC TCGGTCGTTC	5580
GGCTGCGGCG AGCGGTATCA GCTCACTCAA AGGCGGTAAT ACGGTTATCC ACAGAATCAG	5640
GGGATAACGC AGGAAAGAAC ATGTGAGCAA AAGGCCAGCA AAAGGCCAGG AACCGTAAAA	5700
AGGCCGCGTT GCTGGCGTTT TTCCATAGGC TCCGCCCCC TGACGAGCAT CACAAAAATC	5760
GACGCTCAAG TCAGAGGTGG CGAAACCCGA CAGGACTATA AAGATACCAG GCGTTTCCCC	5820
CTGGAAGCTC CCTCGTGCGC TCTCCTGTTC CGACCCTGCC GCTTACCGGA TACCTGTCCG	5880
CCTTTCTCCC TTCGGGAAGC GTGGCGCTTT CTCAATGCTC ACGCTGTAGG TATCTCAGTT	5940
CGGTGTAGGT CGTTGCTCC AAGCTGGGCT GTGTGCACGA ACCCCCCGTT CAGCCCGACC	6000
GCTGCGCCTT ATCCGGTAAC TATCGTCTTG AGTCCAACCC GGTAAGACAC GACTTATCGC	6060
CACTGGCAGC AGCCACTGGT AACAGGATTA GCAGAGCGAG GTATGTAGGC GGTGCTACAG	6120
AGTTCTTGAA GTGGTGGCCT AACTACGGCT AACTAGAAG GACAGTATTT GGTATCTGCG	6180
CTCTGCTGAA GCCAGTTACC TTGGGAAAAA GAGTTGGTAG CTCTTGATCC GGCAAACAAA	6240
CCACCGCTGG TAGCGGTGGT TTTTTTGTTT GCAAGCAGCA GATTACGCGC AGAAAAAAG	6300
GATCTCAAGA AGATCCTTTG ATCTTTTCTA CGGGGTCTGA CGCTCAGTGG AACGAAAAC	6360
CACGTTAAGG GATTTTGGTC ATGAGATTAT CAAAAAGGAT CTTACCTAG ATCCTTTTAA	6420
ATTAAAAATG AAGTTTTAAA TCAATCTAAA GTATATATGA GTAAACTTGG TCTGACAGTT	6480
ACCAATGCTT AATCAGTGAG GCACCTATCT CAGCGATCTG TCTATTCGT TCATCCATAG	6540

TTGCCTGACT CCCCCTCGTG TAGATAACTA CGATACGGGA GGGCTTACCA TCTGGCCCCA	6600
GTGCTGCAAT GATACCGCGA GACCCACGCT CACCGGCTCC AGATTTATCA GCAATAAACC	6660
AGCCAGCCGG AAGGGCCGAG CGCAGAAAGTG GTCCTGCAAC TTTATCCGCC TCCATCCAGT	6720
CTATTAATTG TTGCCGGGAA GCTAGAGTAA GTAGTTCGCC AGTTAATAGT TTGCGCAACG	6780
TTGTTGCCAT TGCTACAGGC ATCGTGGTGT CACGCTCGTC GTTTGGTATG GCTTCATTCA	6840
GCTCCGGTTC CCAACGATCA AGGCGAGTTA CATGATCCCC CATGTTGTGC AAAAAAGCGG	6900
TTAGTCCTT CGGTCCTCCG ATCGTTGTCA GAAGTAAGTT GGCCGCAGTG TTATCACTCA	6960
TGGTTATGGC AGCACTGCAT AATTCTCTTA CTGTCATGCC ATCCGTAAGA TGCTTTTCTG	7020
TGACTGGTGA GTACTCAACC AAGTCATTCT GAGAATAGTG TATGCGGCGA CCGAGTTGCT	7080
CTTGCCCGGC GTCAATACGG GATAATACCG CGCCACATAG CAGAACTTTA AAAGTGCTCA	7140
TCATTGGAAC ACGTTCTTCG GGGCGAAAAC TCTCAAGGAT CTTACCGCTG TTGAGATCCA	7200
GTTTCGATGA ACCCACTCGT GCACCCAACT GATCTTCAGC ATCTTTTACT TTCACCAGCG	7260
TTTCTGGGTG AGCAAAAACA GGAAGGCAAA ATGCCGCAAA AAAGGGAATA AGGGCGACAC	7320
GGAAATGTTG AATACTCATA CTCTTCCTTT TTCAATATTA TTGAAGCATT TATCAGGGTT	7380
ATTGTCTCAT GAGCGGATAC ATATTTGAAT GTATTTAGAA AAATAAACAA ATAGGGGTTC	7440
CGCGCACATT TCCCCGAAAA GTGCCACCTG ACGTC	7475

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8192 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

GACGGATCGG GAGATCTCCC GATCCCCTAT GGTGACTCT CAGTACAATC TGCTCTGATG	60
CCGCATAGTT AAGCCAGTAT CTGCTCCCTG CTTGTGTGTT GGAGGTCGCT GAGTAGTGCG	120
CGAGCAAAAT TTAAGCTACA ACAAGGCAAG GCTTGACCGA CAATTGCATG AAGAATCTGC	180
TTAGGGTTAG GCGTTTTGCG CTGCTTCGCG ATGTACGGGC CAGATATACG CGTTGACATT	240
GATTATTGAC TAGTTATTAA TAGTAATCAA TTACGGGGTC ATTAGTTCAT AGCCCATATA	300

TGGAGTTCCG	CGTTACATAA	CTTACGGTAA	ATGGCCCCGCC	TGGCTGACCG	CCCAACGACC	360
CCCGCCCAT	GACGTCAATA	ATGACGTATG	TTCCCATAGT	AACGCCAATA	GGGACTTTCC	420
ATTGACGTCA	ATGGGTGGAC	TATTTACGGT	AAACTGCCCCA	CTTGGCAGTA	CATCAAGTGT	480
ATCATATGCC	AAGTACGCCC	CCTATTGACG	TCAATGACGG	TAAATGGCCC	GCCTGGCATT	540
ATGCCCAGTA	CATGACCTTA	TGGGACTTTC	CTACTTGGCA	GTACATCTAC	GTATTAGTCA	600
TCGCTATTAC	CATGGTGATG	CGGTTTTGGC	AGTACATCAA	TGGGCGTGGA	TAGCGGTTTG	660
ACTCACGGGG	ATTTCCAAGT	CTCCACCCCA	TTGACGTCAA	TGGGAGTTTG	TTTTGGCACC	720
AAAATCAACG	GGACTTTCCA	AAATGTCGTA	ACAACCTCCGC	CCCATTGACG	CAAATGGGCG	780
GTAGGCGTGT	ACGGTGGGAG	GTCTATATAA	GCAGAGCTCT	CTGGCTAACT	AGAGAACCCA	840
CTGCTTACTG	GCTTATCGAA	ATTAATACGA	CTCACTATAG	GGAGACCCAA	GCTGGCTAGC	900
GTTTAAACGG	GCCCTCTAGA	CTCGAGCGGC	CGCCACTGTG	CTGGATATCT	GCAGAATTCG	960
GCTTGGGATG	ACGCCTCCTC	CGCCCGGACG	TGCCGCCCCC	AGCGCACCGC	GCGCCCGCGT	1020
CCCTGGCCCCG	CCGGCTCGGT	TGGGGCTTCC	GCTGCGGCTG	CGGCTGCTGC	TGCTGCTCTG	1080
GGCGGCCGCC	GCCTCCGCCC	AGGGCCACCT	AAGGAGCGGA	CCCCGCATCT	TCGCCGTCTG	1140
GAAAGGCCAT	GTAGGGCAGG	ACCGGGTGGA	CTTTGGCCAG	ACTGAGCCGC	ACACGGTGCT	1200
TTTCCACGAG	CCAGGCAGCT	CCTCTGTGTG	GGTGGGAGGA	CGTGGCAAGG	TCTACCTCTT	1260
TGACTTCCCC	GAGGGCAAGA	ACGCATCTGT	GCGCACGGTG	AATATCGGCT	CCACAAAGGG	1320
GTCCTGTCTG	GATAAGCGGG	ACTGCGAGAA	CTACATCACT	CTCCTGGAGA	GGCGGAGTGA	1380
GGGGCTGCTG	GCCTGTGGCA	CCAACGCCCCG	GCACCCCAGC	TGCTGGAACC	TGGTGAATGG	1440
CACTGTGGTG	CCACTTGGCG	AGATGAGAGG	CTACGCCCCC	TTCAGCCCCG	ACGAGAACTC	1500
CCTGGTTCTG	TTTGAAGGGG	ACGAGGTGTA	TTCCACCATC	CGBAAGCAGG	AATACAATGG	1560
GAAGATCCCT	CGGTTCCGCC	GCATCCGGGG	CGAGAGTGAG	CTGTACACCA	GTGATACTGT	1620
CATGCAGAAC	CCACAGTTCA	TCAAAGCCAC	CATCGTGCAC	CAAGACCAGG	CTTACGATGA	1680
CAAGATCTAC	TACTTCTTCC	GAGAGGACAA	TCCTGACAAG	AATCCTGAGG	CTCCTCTCAA	1740
TGTGTCCCGT	GTGGCCCAGT	TGTGCAGGGG	GGACCAGGGT	GGGGAAAGTT	CACTGTCACT	1800
CTCCAAGTGG	AACACTTTTC	TGAAAGCCAT	GCTGGTATGC	AGTGATGCTG	CCACCAACAA	1860
GAACCTCAAC	AGGCTGCAAG	ACGTCTTCCT	GCTCCCTGAC	CCCAGCGGCC	AGTGGAGGGA	1920
CACCAGGGTC	TATGGTGTTT	TCTCCAACCC	CTGGAACCTAC	TCAGCCGTCT	GTGTGTATTG	1980
CCTCGGTGAC	ATTGACAAGG	TCTTCCGTAC	CTCCTCACTC	AAGGGCTACC	ACTCAAGCCT	2040

TCCCAACCCG	CGGCCTGGCA	AGTGCCTCCC	AGACCAGCAG	CCGATACCCA	CAGAGACCTT	2100
CCAGGTGGCT	GACCGTCACC	CAGAGGTGGC	GCAGAGGGTG	GAGCCCATGG	GGCCTCTGAA	2160
GACGCCATTG	TTCCACTCTA	AATACCACTA	CCAGAAAGTG	GCCGTTCACC	GCATGCAAGC	2220
CAGCCACGGG	GAGACCTTTC	ATGTGCTTTA	CCTAACTACA	GACAGGGGCA	CTATCCACAA	2280
GGTGGTGGAA	CCGGGGGAGC	AGGAGCACAG	CTTCGCCTTC	AACATCATGG	AGATCCAGCC	2340
CTTCCGCCGC	GCGGCTGCCA	TCCAGACCAT	GTCGCTGGAT	GCTGAGCGGA	GGAAGCTGTA	2400
TGTGAGCTCC	CAGTGGGAGG	TGAGCCAGGT	GCCCCTGGAC	CTGTGTGAGG	TCTATGGCGG	2460
GGGCTGCCAC	GGTTGCCTCA	TGTCCCGAGA	CCCCTACTGC	GGCTGGGACC	AGGGCCGCTG	2520
CATCTCCATC	TACAGCTCCG	AACGGTCAGT	GCTGCAATCC	ATTAATCCAG	CCGAGCCACA	2580
CAAGGAGTGT	CCCAACCCCA	AACCAGACAA	GGCCCCACTG	CAGAAGGTTT	CCCTGGCCCC	2640
AAACTCTCGC	TACTACCTGA	GCTGCCCCAT	GGAATCCCGC	CACGCCACCT	ACTCATGGCG	2700
CCACAAGGAG	AACGTGGAGC	AGAGCTGCGA	ACCTGGTCAC	CAGAGCCCCA	ACTGCATCCT	2760
GTTTCATCGAG	AACCTCACGG	CGCAGCAGTA	CGGCCACTAC	TTCTGCGAGG	CCCAGGAGGG	2820
CTCCTACTTC	CGCGAGGCTC	AGCACTGGCA	GCTGCTGCCC	GAGGACGGCA	TCATGGCCGA	2880
GCACCTGCTG	GGTCATGCCT	GTGCCCTGGC	TGCCTCCCTC	TGGCTGGGGG	TGCTGCCCAC	2940
ACTCACTCTT	GGCTTGCTGG	TCCACATGGT	GAGCAAGGGC	GAGGAGCTGT	TCACCGGGGT	3000
GGTGCCCATC	CTGGTCGAGC	TGGACGGCGA	CGTAAACGGC	CACAAGTTCA	GCGTGTCCGG	3060
CGAGGGCGAG	GGCGATGCCA	CCTACGGCAA	GCTGACCCTG	AAGTTCATCT	GCACCACCGG	3120
CAAGCTGCCC	GTGCCCTGGC	CCACCCTCGT	GACCACCCTG	ACCTACGGCG	TGCAGTGCTT	3180
CAGCCGCTAC	CCCGACCACA	TGAAGCAGCA	CGACTTCTTC	AAGTCCGCCA	TGCCCCGAAGG	3240
CTACGTCCAG	GAGCGCACCA	TCTTCTTCAA	GGACGACGGC	AACTACAAGA	CCCGCGCCGA	3300
GGTGAAGTTC	GAGGGCGACA	CCCTGGTGAA	CCGCATCGAG	CTGAAGGGCA	TCGACTTCAA	3360
GGAGGACGGC	AACATCCTGG	GGCACAAGCT	GGAGTACAAC	TACAACAGCC	ACAACGTCTA	3420
TATCATGGCC	GACAAGCAGA	AGAACGGCAT	CAAGGTGAAC	TTCAAGATCC	GCCACAACAT	3480
CGAGGACGGC	AGCGTGCAGC	TCGCCGACCA	CTACCAGCAG	AACACCCCCA	TCGGCGACGG	3540
CCCCGTGCTG	CTGCCCCACA	ACCACTACCT	GAGCACCCAG	TCCGCCCTGA	GCAAAGACCC	3600
CAACGAGAAG	CGCGATCACA	TGGTCCTGCT	GGAGTTCGTG	ACCGCCGCCG	GGATCACTCT	3660
CGGCATGGAC	GAGCTGTACA	AGGTGAAGCT	TGGGCCCCGA	CAAAAACCTCA	TCTCAGAAGA	3720

GGATCTGAAT	AGCGCCGTCG	ACCATCATCA	TCATCATCAT	TGAGTTTAAA	CCGCTGATCA	3780
GCCTCGACTG	TGCCTTCTAG	TTGCCAGCCA	TCTGTTGTTT	GCCCCTCCCC	CGTGCCTTCC	3840
TTGACCCTGG	AAGGTGCCAC	TCCCAGTGTC	CTTTCCTAAT	AAAATGAGGA	AATTGCATCG	3900
CATTGTCTGA	GTAGGTGTCA	TTCTATTCTG	GGGGGTGGGG	TGGGGCAGGA	CAGCAAGGGG	3960
GAGGATTGGG	AAGACAATAG	CAGGCATGCT	GGGGATGCGG	TGGGCTCTAT	GGCTTCTGAG	4020
GCGGAAAGAA	CCAGCTGGGG	CTCTAGGGGG	TATCCCCACG	CGCCCTGTAG	CGGCGCATTA	4080
AGCGCGGCGG	GTGTGGTGGT	TACGCGCAGC	GTGACCGCTA	CACTTGCCAG	CGCCCTAGCG	4140
CCCGCTCCTT	TCGCTTTCTT	CCCTTCCTTT	CTCGCCACGT	TCGCCGGCTT	TCCCCGTCAA	4200
GCTCTAAATC	GGGGCATCCC	TTTAGGGTTC	CGATTTAGTG	CTTTACGGCA	CCTCGACCCC	4260
AAAAAACTTG	ATTAGGGTGA	TGGTTCACGT	AGTGGGCCAT	CGCCCTGATA	GACGGTTTTT	4320
CGCCCTTTGA	CGTTGGAGTC	CACGTTCTTT	AATAGTGGAC	TCTTGTTCCA	AACTGGAACA	4380
AACTCAACC	CTATCTCGGT	CTATTCTTTT	GATTTATAAG	GGATTTTGGG	GATTTTCGGC	4440
TATTGGTTAA	AAAATGAGCT	GATTTAACAA	AAATTTAACG	CGAATTAATT	CTGTGGAATG	4500
TGTGTCAGTT	AGGGTGTGGA	AAGTCCCCAG	GCTCCCCAGG	CAGGCAGAAG	TATGCAAAGC	4560
ATGCATCTCA	ATTAGTCAGC	AACCAGGTGT	GGAAAGTCCC	CAGGCTCCCC	AGCAGGCAGA	4620
AGTATGCAAA	GCATGCATCT	CAATTAGTCA	GCAACCATAG	TCCCGCCCCT	AACTCCGCCC	4680
ATCCCGCCCC	TAACTCCGCC	CAGTTCCGCC	CATTCTCCGC	CCCATGGCTG	ACTAATTTTT	4740
TTTATTTATG	CAGAGGCCGA	GGCCGCCTCT	GCCTCTGAGC	TATTCCAGAA	GTAGTGAGGA	4800
GGCTTTTTTG	GAGGCCTAGG	CTTTTGCAAA	AAGCTCCCGG	GAGCTTGTAT	ATCCATTTTC	4860
GGATCTGATC	AAGAGACAGG	ATGAGGATCG	TTTCGCATGA	TTGAACAAGA	TGGATTGCAC	4920
GCAGGTTCTC	CGGCCGCTTG	GGTGGAGAGG	CTATTCGGCT	ATGACTGGGC	ACAACAGACA	4980
ATCGGCTGCT	CTGATGCCGC	CGTGTTCCGG	CTGTCAGCGC	AGGGGCGCCC	GGTTCTTTTT	5040
GTCAAGACCG	ACCTGTCCGG	TGCCCTGAAT	GAACTGCAGG	ACGAGGCAGC	GCGGCTATCG	5100
TGGCTGGCCA	CGACGGGCGT	TCCTTGCGCA	GCTGTGCTCG	ACGTTGTCAC	TGAAGCGGGA	5160
AGGGACTGGC	TGCTATTGGG	CGAAGTGCCG	GGGCAGGATC	TCCTGTCATC	TCACCTTGCT	5220
CCTGCCGAGA	AAGTATCCAT	CATGGCTGAT	GCAATGCGGC	GGCTGCATAC	GCTTGATCCG	5280
GCTACCTGCC	CATTCGACCA	CCAAGCGAAA	CATCGCATCG	AGCGAGCACG	TACTCGGATG	5340
GAAGCCGGTC	TTGTGATCA	GGATGATCTG	GACGAAGAGC	ATCAGGGGCT	CGCGCCAGCC	5400
GAACTGTTCG	CCAGGCTCAA	GGCGCGCATG	CCCGACGGCG	AGGATCTCGT	CGTGACCCAT	5460

GGCGATGCCCT GCTTGCCGAA TATCATGGTG GAAAATGGCC GCTTTTCTGG ATTCATCGAC	5520
TGTGGCCGGC TGGGTGTGGC GGACCGCTAT CAGGACATAG CGTTGGCTAC CCGTGATATT	5580
GCTGAAGAGC TTGGCGGCGA ATGGGCTGAC CGCTTCCTCG TGCTTTACGG TATCGCCGCT	5640
CCCGATTTCG AGCGCATCGC CTTCTATCGC CTTCTTGACG AGTTCTTCTG AGCGGGACTC	5700
TGGGGTTTCGA AATGACCGAC CAAGCGACGC CCAACCTGCC ATCACGAGAT TTCGATTCCA	5760
CCGCCGCCTT CTATGAAAGG TTGGGCTTCG GAATCGTTTT CCGGGACGCC GGCTGGATGA	5820
TCCTCCAGCG CGGGGATCTC ATGCTGGAGT TCTTCGCCCC CCCCAACTTG TTTATTGCAG	5880
CTTATAATGG TTACAAATAA AGCAATAGCA TCACAAATTT CACAAATAAA GCATTTTTTTT	5940
CACTGCATTC TAGTTGTGGT TTGTCCAAAC TCATCAATGT ATCTTATCAT GTCTGTATAC	6000
CGTCGACCTC TAGCTAGAGC TTGGCGTAAT CATGGTCATA GCTGTTTCCT GTGTGAAATT	6060
GTTATCCGCT CACAATTCCA CACAACATAC GAGCCGGAAG CATAAAGTGT AAAGCCTGGG	6120
GTGCCTAATG AGTGAGCTAA CTCACATTAA TTGCGTTGCG CTCACTGCCC GCTTTCAGT	6180
CGGGAAACCT GTCGTGCCAG CTGCATTAAT GAATCGGCCA ACGCGCGGGG AGAGGCGGTT	6240
TGCGTATTGG GCGCTCTTCC GCTTCCTCGC TCACTGACTC GCTGCGCTCG GTCGTTCCGC	6300
TGCGGCGAGC GGTATCAGCT CACTCAAAGG CGGTAATACG GTTATCCACA GAATCAGGGG	6360
ATAACGCAGG AAAGAACATG TGAGCAAAAG GCCAGCAAAA GGCCAGGAAC CGTAAAAAGG	6420
CCGCGTTGCT GCGTTTTTTC CATAGGCTCC GCCCCCTGA CGAGCATCAC AAAAATCGAC	6480
GCTCAAGTCA GAGGTGGCGA AACCCGACAG GACTATAAAG ATACCAGGCG TTTCCCCCTG	6540
GAAGCTCCCT CGTGCGCTCT CCTGTTCCGA CCCTGCCGCT TACCGGATAC CTGTCCGCCT	6600
TTCTCCCTTC GGAAGCGTG GCGCTTTCTC AATGCTCACG CTGTAGGTAT CTCAGTTCGG	6660
TGTAGGTCGT TCGCTCCAAG CTGGGCTGTG TGCACGAACC CCCCGTTCAG CCCGACCGCT	6720
GCGCCTTATC CGGTAACAT CGTCTTGAGT CCAACCCGGT AAGACACGAC TTATCGCCAC	6780
TGGCAGCAGC CACTGGTAAC AGGATTAGCA GAGCGAGGTA TGTAGGCGGT GCTACAGAGT	6840
TCTTGAAAGTG GTGGCCTAAC TACGGCTACA CTAGAAGGAC AGTATTTGGT ATCTGCGCTC	6900
TGCTGAAGCC AGTTACCTTC GGAAAAAGAG TTGGTAGCTC TTGATCCGGC AAACAAACCA	6960
CCGCTGGTAG CGGTGGTTTT TTTGTTTGCA AGCAGCAGAT TACGCGCAGA AAAAAAGGAT	7020
CTCAAGAAGA TCCTTTGATC TTTTCTACGG GGTCTGACGC TCAGTGGAAC GAAAATCAC	7080
GTTAAGGGAT TTTGGTCATG AGATTATCAA AAAGGATCTT CACCTAGATC CTTTTAAATT	7140

AAAAATGAAG TTTTAAATCA ATCTAAAGTA TATATGAGTA AACTTGGTCT GACAGTTACC	7200
AATGCTTAAT CAGTGAGGCA CCTATCTCAG CGATCTGTCT ATTTTCGTTCA TCCATAGTTG	7260
CCTGACTCCC CGTCGTGTAG ATAACCTACGA TACGGGAGGG CTTACCATCT GGGCCCACTG	7320
CTGCAATGAT ACCGCGAGAC CCACGCTCAC CGGCTCCAGA TTTATCAGCA ATAAACCAGC	7380
CAGCCGGAAG GGCCGAGCGC AGAAGTGGTC CTGCAACTTT ATCCGCCTCC ATCCAGTCTA	7440
TTAATTGTTG CCGGGAAGCT AGAGTAAGTA GTTCGCCAGT TAATAGTTTG CGCAACGTTG	7500
TTGCCATTGC TACAGGCATC GTGGTGTAC GCTCGTCGTT TGGTATGGCT TCATTCAGCT	7560
C CGGTTCCCA ACGATCAAGG CGAGTTACAT GATCCCCAT GTTGTGCAAA AAAGCGGTTA	7620
GCTCCTTCGG TCCTCCGATC GTTGTACAGAA GTAAGTTGGC CGCAGTGTTA TCACTCATGG	7680
TTATGGCAGC ACTGCATAAT TCTCTTACTG TCATGCCATC CGTAAGATGC TTTTCTGTGA	7740
CTGGTGAGTA CTCAACCAAG TCATTCTGAG AATAGTGAT GCGGCGACCG AGTTGCTCTT	7800
GCCCGGCGTC AATACGGGAT AATACCGCGC CACATAGCAG AACTTTAAAA GTGCTCATCA	7860
TTGGAAAACG TTCTTCGGGG CGAAAATCT CAAGGATCTT ACCGCTGTTG AGATCCAGTT	7920
CGATGTAACC CACTCGTGCA CCAACTGAT CTTTCAGCATC TTTTACTTTC ACCAGCGTTT	7980
CTGGGTGAGC AAAACAGGA AGGCAAAATG CCGCAAAAAA GGAATAAGG GCGACACGGA	8040
AATGTTGAAT ACTCATACTC TTCCTTTTTC AATATTATTG AAGCATTTAT CAGGGTTATT	8100
GTCTCATGAG CGGATACATA TTTGAATGTA TTTAGAAAAA TAAACAAATA GGGGTTCGCG	8160
GCACATTTCC CCGAAAAGTG CCACCTGACG TC	8192

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7000 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

AGATCTCGGC CGCATATTAA GTGCATTGTT CTCGATACCG CTAAGTGCAT TGTTCTCGTT	60
AGCTCGATGG ACAAGTGCAT TGTTCTCTTG CTGAAAGCTC GATGGACAAG TGCATTGTTT	120
TCTTGCTGAA AGCTCGATGG ACAAGTGCAT TGTTCTCTTG CTGAAAGCTC AGTACCCGGG	180

AGTACCCTCG ACCGCCGGAG TATAAATAGA GCGCTTCGT CTACGGAGCG ACAATTCAAT	240
TCAAACAAGC AAAGTGAACA CGTCGCTAAG CGAAAGCTAA GCAAATAAAC AAGCGCAGCT	300
GAACAAGCTA AACAATCTGC AGTAAAGTGC AAGTTAAAGT GAATCAATTA AAAGTAACCA	360
GCAACCAAGT AAATCAACTG CAACTACTGA AATCTGCCAA GAAGTAATTA TTGAATACAA	420
GAAGAGAACT CTGAATACTT TCAACAAGTT ACCGAGAAAG AAGAACTCAC ACACAGCTAG	480
CGTTTAAACT TAAGCTTGGT ACCGAGCTCG GATCCACTAG TCCAGTGTGG TGGAAATTCGG	540
CTTGGGATGA CGCCTCCTCC GCCCGGACGT GCCGCCCCCA GCGCACCGCG CGCCCCGCTC	600
CCTGGCCCCG CGGCTCGGTT GGGGCTTCCG CTGCGGCTGC GGCTGCTGCT GCTGCTCTGG	660
GCGGCCGCCG CCTCCGCCCA GGGCCACCTA AGGAGCGGAC CCCGCATCTT CGCCGTCTGG	720
AAAGGCCATG TAGGGCAGGA CCGGGTGGAC TTTGGCCAGA CTGAGCCGCA CACGGTGCTT	780
TTCCACGAGC CAGGCAGCTC CTCTGTGTGG GTGGGAGGAC GTGGCAAGGT CTACCTCTTT	840
GACTTCCCCG AGGGCAAGAA CGCATCTGTG CGCACGGTGA ATATCGGCTC CACAAAGGGG	900
TCCTGTCTGG ATAAGCGGGA CTGCGAGAAC TACATCACTC TCCTGGAGAG GCGGAGTGAG	960
GGGCTGCTGG CCTGTGGCAC CAACGCCCCG CACCCAGCT GCTGGAACCT GGTGAATGGC	1020
ACTGTGGTGC CACTTGGCGA GATGAGAGGC TACGCCCCCT TCAGCCCGGA CGAGAACTCC	1080
CTGGTTCTGT TTGAAGGGGA CGAGGTGTAT TCCACCATCC GGAAGCAGGA ATACAATGGG	1140
AAGATCCCTC GGTTCCGCCG CATCCGGGGC GAGAGTGAGC TGTACACCAG TGATACTGTC	1200
ATGCAGAACC CACAGTTCAT CAAAGCCACC ATCGTGCACC AAGACCAGGC TTACGATGAC	1260
AAGATCTACT ACTTCTTCCG AGAGGACAAT CCTGACAAGA ATCCTGAGGC TCCTCTCAAT	1320
GTGTCCCGTG TGGCCAGTT GTGCAGGGGG GACCAGGGTG GGGAAAGTTC ACTGTCAGTC	1380
TCCAAGTGGA ACACTTTTCT GAAAGCCATG CTGGTATGCA GTGATGCTGC CACCAACAAG	1440
AACTTCAACA GGCTGCAAGA CGTCTTCCTG CTCCCTGACC CCAGCGGCCA GTGGAGGGAC	1500
ACCAGGGTCT ATGGTGTTTT CTCCAACCCC TGGAACTACT CAGCCGTCTG TGTGTATTCC	1560
CTCGGTGACA TTGACAAGGT CTTCCTGACC TCCTCACTCA AGGGCTACCA CTCAAGCCTT	1620
CCCAACCCGC GGCCTGGCAA GTGCCTCCCA GACCAGCAGC CGATACCCAC AGAGACCTTC	1680
CAGGTGGCTG ACCGTCACCC AGAGGTGGCG CAGAGGGTGG AGCCCATGGG GCCTCTGAAG	1740
ACGCCATTGT TCCACTCTAA ATACCACTAC CAGAAAGTGG CCGTTCACCG CATGCAAGCC	1800
AGCCACGGGG AGACCTTTCA TGTGCTTTAC CTAACTACAG ACAGGGGCAC TATCCACAAG	1860
GTGGTGGAAC CGGGGGAGCA GGAGCACAGC TTCGCCTTCA ACATCATGGA GATCCAGCCC	1920

TTCCGCCGCG	CGGCTGCCAT	CCAGACCATG	TCGCTGGATG	CTGAGCGGAG	GAAGCTGTAT	1980
GTGAGCTCCC	AGTGGGAGGT	GAGCCAGGTG	CCCCTGGACC	TGTGTGAGGT	CTATGGCGGG	2040
GGCTGCCACG	GTTGCCTCAT	GTCCCAGAGAC	CCCTACTGCG	GCTGGGACCA	GGGCCGCTGC	2100
ATCTCCATCT	ACAGCTCCGA	ACGGTCAGTG	CTGCAATCCA	TTAATCCAGC	CGAGCCACAC	2160
AAGGAGTGTC	CCAACCCCAA	ACCAGACAAG	GCCCCACTGC	AGAAGGTTTC	CCTGGCCCCA	2220
AACTCTCGCT	ACTACCTGAG	CTGCCCCATG	GAATCCCGCC	ACGCCACCTA	CTCATGGCGC	2280
CACAAGGAGA	ACGTGGAGCA	GAGCTGCGAA	CCTGGTCACC	AGAGCCCCAA	CTGCATCCTG	2340
TTCATCGAGA	ACCTCACGGC	GCAGCAGTAC	GGCCACTACT	TCTGCGAGGC	CCAGGAGGGC	2400
TCCTACTTCC	GCGAGGCTCA	GCACTGGCAG	CTGCTGCCCC	AGGACGGCAT	CATGGCCGAG	2460
CACCTGCTGG	GTCATGCCTG	TGCCCTGGCT	GCCTCCCTCT	GGCTGGGGGT	GCTGCCCACA	2520
CTCACTCTTG	GCTTGCTGGT	CCACGTGAAG	CTTGGGCCCC	TTTAAACCCG	CTGATCAGCC	2580
TCGACTGTGC	CTTCTAGTTG	CCAGCCATCT	GTGTTTGCC	CCTCCCCCGT	GCCTTCCTTG	2640
ACCCTGGAAG	GTGCCACTCC	CACTGTCCTT	TCCTAATAAA	ATGAGGAAAT	TGCATCGCAT	2700
TGTCTGAGTA	GGTGTCAATC	TATTCTGGGG	GGTGGGGTGG	GGCAGGACAG	CAAGGGGGAG	2760
GATTGGGAAG	ACAATAGCAG	GCATGCTGGG	GATGCGGTGG	GCTCTATGGC	TTCTGAGGCG	2820
GAAAGAACCA	GCTGGGGCTC	TAGGGGGTAT	CCCCACGCGC	CCTGTAGCGG	CGCATTAAGC	2880
GCGGCGGGTG	TGGTGGTTAC	GCGCAGCGTG	ACCCTACAC	TTGCCAGCGC	CCTAGCGCCC	2940
GCTCCTTTTCG	CTTTCTTCCC	TTCCTTTCTC	GCCACGTTTCG	CCGGCTTTCC	CCGTCAAGCT	3000
CTAAATCGGG	GCATCCCTTT	AGGGTTCCGA	TTTAGTGCTT	TACGGCACCT	CGACCCCAAA	3060
AAACTTGATT	AGGGTGATGG	TTCACGTAGT	GGGCCATCGC	CCTGATAGAC	GGTTTTTCGC	3120
CCTTTGACGT	TGGAGTCCAC	GTTCTTTAAT	AGTGGACTCT	TGTTCCAAAC	TGGAACAACA	3180
CTCAACCCTA	TCTCGGTCTA	TTCTTTTGAT	TTATAAGGGA	TTTGGGGAT	TTCGGCCTAT	3240
TGGTTAAAAA	ATGAGCTGAT	TTAACAAAAA	TTAACGCGA	ATTAATTCTG	TGGAATGTGT	3300
GTCAGTTAGG	GTGTGGAAAAG	TCCCCAGGCT	CCCCAGGCAG	GCAGAAGTAT	GCAAAGCATG	3360
CATCTCAATT	AGTCAGCAAC	CAGGTGTGGA	AAGTCCCCAG	GCTCCCCAGC	AGGCAGAAGT	3420
ATGCAAAGCA	TGCATCTCAA	TTAGTCAGCA	ACCATAGTCC	CGCCCCTAAC	TCCGCCCATC	3480
CCGCCCCCTAA	CTCCGCCCCAG	TTCCGCCCAT	TCTCCGCCCC	ATGGCTGACT	AATTTTTTTT	3540
ATTTATGCAG	AGGCCGAGGC	CGCCTCTGCC	TCTGAGCTAT	TCCAGAAGTA	GTGAGGAGGC	3600

TTTTTTGGAG GCCTAGGCTT TTGCAAAAAG CTCCCGGGAG CTTGTATATC CATTTTCGGA	3660
TCTGATCAAG AGACAGGATG AGGATCGTTT CGCATGATTG AACAAGATGG ATTGCACGCA	3720
GGTTCTCCGG CCGCTTGGGT GGAGAGGCTA TTCGGCTATG ACTGGGCACA ACAGACAATC	3780
GGCTGCTCTG ATGCCGCCGT GTTCCGGCTG TCAGCGCAGG GGCGCCCGGT TCTTTTTGTC	3840
AAGACCGACC TGTCCGGTGC CCTGAATGAA CTGCAGGACG AGGCAGCGCG GCTATCGTGG	3900
CTGGCCACGA CGGGCGTTCC TTGCGCAGCT GTGCTCGACG TTGTCACTGA AGCGGGAAGG	3960
GACTGGCTGC TATTGGGCGA AGTGCCGGGG CAGGATCTCC TGTCACTCA CCTTGCTCCT	4020
GCCGAGAAAG TATCCATCAT GGCTGATGCA ATGCGGCGGC TGCATACGCT TGATCCGGCT	4080
ACCTGCCCAT TCGACCACCA AGCGAAACAT CGCATCGAGC GAGCACGTAC TCGGATGGAA	4140
GCCGGTCTTG TCGATCAGGA TGATCTGGAC GAAGAGCATC AGGGGCTCGC GCCAGCCGAA	4200
CTGTTTCGCCA GGCTCAAGGC GCGCATGCCC GACGGCGAGG ATCTCGTCGT GACCCATGGC	4260
GATGCCTGCT TGCCGAATAT CATGGTGGAA AATGGCCGCT TTTCTGGATT CATCGACTGT	4320
GGCCGGCTGG GTGTGGCGGA CCGCTATCAG GACATAGCGT TGGCTACCCG TGATATTGCT	4380
GAAGAGCTTG GCGGCGAATG GGCTGACCGC TTCCTCGTGC TTTACGGTAT CGCCGCTCCC	4440
GATTCGCAGC GCATCGCCTT CTATCGCCTT CTTGACGAGT TCTTCTGAGC GGGACTCTGG	4500
GGTTCGAAAT GACCGACCAA GCGACGCCCA ACCTGCCATC ACAGATTTT GATTCCACCG	4560
CCGCCTTCTA TGAAAGGTTG GGCTTCGGAA TCGTTTTCCG GGACGCCGGC TGGATGATCC	4620
TCCAGCGCGG GGATCTCATG CTGGAGTTCT TCGCCACCCC CAACTTGTTT ATTGCAGCTT	4680
ATAATGGTTA CAAATAAAGC AATAGCATCA CAAATTTTAC AAATAAAGCA TTTTTTTTAC	4740
TGCATTCTAG TTGTGGTTTG TCCAAACTCA TCAATGTATC TTATCATGTC TGTATACCGT	4800
CGACCTCTAG CTAGAGCTTG GCGTAATCAT GGTCATAGCT GTTTCCTGTG TGAAATTGTT	4860
ATCCGCTCAC AATTCCACAC AACATACGAG CCGGAAGCAT AAAGTGTAAG GCCTGGGGTG	4920
CCTAATGAGT GAGCTAACTC ACATTAATTG CGTTGCGCTC ACTGCCCGCT TTCCAGTCGG	4980
GAAACCTGTC GTGCCAGCTG CATTAATGAA TCGGCCAACG CGCGGGGAGA GGCGGTTTGC	5040
GTATTGGGCG CTCTTCCGCT TCCTCGCTCA CTGACTCGCT GCGCTCGGTC GTTCGGCTGC	5100
GGCGAGCGGT ATCAGCTCAC TCAAAGGCGG TAATACGGTT ATCCACAGAA TCAGGGGATA	5160
ACGCAGGAAA GAACATGTGA GCAAAAGGCC AGCAAAAGGC CAGGAACCGT AAAAAGGCCG	5220
CGTTGCTGGC GTTTTTCCAT AGGCTCCGCC CCCCTGACGA GCATCACAAA AATCGACGCT	5280
CAAGTCAGAG GTGGCGAAAC CCGACAGGAC TATAAAGATA CCAGGCGTTT CCCCCTGGAA	5340

GCTCCCTCGT	GCGCTCTCCT	GTTCCGACCC	TGCCGCTTAC	CGGATACCTG	TCCGCCTTTC	5400
TCCCTTCGGG	AAGCGTGGCG	CTTTCTCAAT	GCTCACGCTG	TAGGTATCTC	AGTTCGGTGT	5460
AGGTCGTTTC	CTCCAAGCTG	GGCTGTGTGC	ACGAACCCCC	CGTTCAGCCC	GACCGCTGCG	5520
CCTTATCCGG	TAACTATCGT	CTTGAGTCCA	ACCCGGTAAG	ACACGACTTA	TCGCCACTGG	5580
CAGCAGCCAC	TGGTAACAGG	ATTAGCAGAG	CGAGGTATGT	AGGCGGTGCT	ACAGAGTTCT	5640
TGAAGTGGTG	GCCTAACTAC	GGCTACACTA	GAAGGACAGT	ATTTGGTATC	TGCGCTCTGC	5700
TGAAGCCAGT	TACCTTCGGA	AAAAGAGTTG	GTAGCTCTTG	ATCCGGCAAA	CAAACCACCG	5760
CTGGTAGCGG	TGGTTTTTTT	GTTTGCAAGC	AGCAGATTAC	GCGCAGAAAA	AAAGGATCTC	5820
AAGAAGATCC	TTTGATCTTT	TCTACGGGGT	CTGACGCTCA	GTGGAACGAA	AACTCACGTT	5880
AAGGGATTTT	GGTCATGAGA	TTATCAAAAA	GGATCTTCAC	CTAGATCCTT	TTAAATTAAA	5940
AATGAAGTTT	TAAATCAATC	TAAAGTATAT	ATGAGTAAAC	TTGGTCTGAC	AGTTACCAAT	6000
GCTTAATCAG	TGAGGCACCT	ATCTCAGCGA	TCTGTCTATT	TCGTTCATCC	ATAGTTGCCT	6060
GACTCCCCGT	CGTGTAGATA	ACTACGATAC	GGGAGGGCTT	ACCATCTGGC	CCCAGTGCTG	6120
CAATGATACC	GCGAGACCCA	CGCTCACCGG	CTCCAGATTT	ATCAGCAATA	AACCAGCCAG	6180
CCGGAAGGGC	CGAGCGCAGA	AGTGGTCCTG	CAACTTTATC	CGCCTCCATC	CAGTCTATTA	6240
ATTGTTGCCG	GGAAGCTAGA	GTAAGTAGTT	CGCCAGTTAA	TAGTTTGCGC	AACGTTGTTG	6300
CCATTGCTAC	AGGCATCGTG	GTGTCACGCT	CGTCGTTTGG	TATGGCTTCA	TTCAGCTCCG	6360
GTTCCCAACG	ATCAAGGCGA	GTTACATGAT	CCCCCATGTT	GTGCAAAAAA	GCGGTTAGCT	6420
CCTTCGGTCC	TCCGATCGTT	GTCAGAAGTA	AGTTGGCCGC	AGTGTTATCA	CTCATGGTTA	6480
TGGCAGCACT	GCATAATTCT	CTTACTGTCA	TGCCATCCGT	AAGATGCTTT	TCTGTGACTG	6540
GTGAGTACTC	AACCAAGTCA	TTCTGAGAAT	AGTGTATGCG	GCGACCGAGT	TGCTCTTGCC	6600
CGGCGTCAAT	ACGGGATAAT	ACCGCGCCAC	ATAGCAGAAC	TTTAAAAGTG	CTCATCATTG	6660
GAAAACGTTT	TTCGGGGCGA	AAACTCTCAA	GGATCTTACC	GCTGTTGAGA	TCCAGTTCGA	6720
TGTAACCCAC	TCGTGCACCC	AACTGATCTT	CAGCATCTTT	TACTTTCACC	AGCGTTTCTG	6780
GGTGAGCAAA	AACAGGAAGG	CAAAATGCCG	CAAAAAAGGG	AATAAGGGCG	ACACGGAAAT	6840
GTTGAATACT	CATACTCTTC	CTTTTTCAAT	ATTATTGAAG	CATTTATCAG	GGTTATTGTC	6900
TCATGAGCGG	ATACATATTT	GAATGTATTT	AGAAAAATAA	ACAAATAGGG	GTTCCGCGCA	6960
CATTTCCCCG	AAAAGTGCCA	CCTGACGTCG	ACGGATCGGG			7000

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7108 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

AGATCTCGGC CGCATATTAA GTGCATTGTT CTCGATACCG CTAAGTGCAT TGTTCTCGTT	60
AGCTCGATGG ACAAGTGCAT TGTTCTCTTG CTGAAAGCTC GATGGACAAG TGCATTGTTC	120
TCTTGCTGAA AGCTCGATGG ACAAGTGCAT TGTTCTCTTG CTGAAAGCTC AGTACCCGGG	180
AGTACCCTCG ACCGCCGGAG TATAAATAGA GGCCTTCGT CTACGGAGCG ACAATTCAAT	240
TCAAACAAGC AAAGTGAACA CGTCGCTAAG CGAAAGCTAA GCAAATAAAC AAGCGCAGCT	300
GAACAAGCTA AACAACTCTGC AGTAAAGTGC AAGTTAAAGT GAATCAATTA AAAGTAACCA	360
GCAACCAAGT AAATCAACTG CAACTACTGA AATCTGCCAA GAAGTAATTA TTGAATACAA	420
GAAGAGAACT CTGAATACTT TCAACAAGTT ACCGAGAAAAG AAGAACTCAC ACACAGCTAG	480
CGTTTAAACT TAAGCTTGGT ACCGAGCTCG GATCCACTAG TCCAGTGTGG TGGAATTCGG	540
CTTGGGATGA CGCCTCCTCC GCCCGACGT GCCGCCCCCA GCGCACCGCG CGCCCGCGTC	600
CCTGGCCCCG CGGCTCGGTT GGGGCTTCCG CTGCGGCTGC GGCTGCTGCT GCTGCTCTGG	660
GCGGCCGCCG CCTCCGCCCA GGGCCACCTA AGGAGCGGAC CCCGCATCTT CGCCGTCTGG	720
AAAGGCCATG TAGGGCAGGA CCGGGTGGAC TTTGGCCAGA CTGAGCCGCA CACGGTGCTT	780
TTCCACGAGC CAGGCAGCTC CTCTGTGTGG GTGGGAGGAC GTGGCAAGGT CTACCTCTTT	840
GACTTCCCCG AGGGCAAGAA CGCATCTGTG CGCACGGTGA ATATCGGCTC CACAAAGGGG	900
TCCTGTCTGG ATAAGCGGGA CTGCGAGAAC TACATCACTC TCCTGGAGAG GCGGAGTGAG	960
GGGCTGCTGG CCTGTGGCAC CAACGCCCCG CACCCCAGCT GCTGGAACCT GGTGAATGGC	1020
ACTGTGGTGC CACTTGGCGA GATGAGAGGC TACGCCCCCT TCAGCCCGGA CGAGAACTCC	1080
CTGGTTCTGT TTGAAGGGGA CGAGGTGTAT TCCACCATCC GGAAGCAGGA ATACAATGGG	1140
AAGATCCCTC GGTTCGCGG CATCCGGGGC GAGAGTGAGC TGTACACCAG TGATACTGTC	1200
ATGCAGAACC CACAGTTCAT CAAAGCCACC ATCGTGCACC AAGACCAGGC TTACGATGAC	1260

AAGATCTACT ACTTCTTCCG AGAGGACAAT CCTGACAAGA ATCCTGAGGC TCCTCTCAAT	1320
GTGTCCCGTG TGGCCAGTT GTGCAGGGGG GACCAGGGTG GGGAAAGTTC ACTGTCAGTC	1380
TCCAAGTGGA ACACTTTTCT GAAAGCCATG CTGGTATGCA GTGATGCTGC CACCAACAAG	1440
AACTTCAACA GGCTGCAAGA CGTCTTCTG CTCCCTGACC CCAGCGGCCA GTGGAGGGAC	1500
ACCAGGGTCT ATGGTGTTTT CTCCAACCCC TGGAAGTACT CAGCCGTCTG TGTGTATTCC	1560
CTCGGTGACA TTGACAAGGT CTTCCGTACC TCCTCACTCA AGGGCTACCA CTCAAGCCTT	1620
CCCAACCCGC GGCCTGGCAA GTGCCTCCCA GACCAGCAGC CGATACCCAC AGAGACCTTC	1680
CAGGTGGCTG ACCGTCACCC AGAGGTGGCG CAGAGGGTGG AGCCCATGGG GCCTCTGAAG	1740
ACGCCATTGT TCCACTCTAA ATACCACTAC CAGAAAGTGG CCGTTCACCG CATGCAAGCC	1800
AGCCACGGGG AGACCTTTCA TGTGCTTTAC CTAAGTACAG ACAGGGGCAC TATCCACAAG	1860
GTGGTGGAAC CGGGGGAGCA GGAGCACAGC TTCGCCTTCA ACATCATGGA GATCCAGCCC	1920
TTCCGCCGCG CGGCTGCCAT CCAGACCATG TCGCTGGATG CTGAGCGGAG GAAGCTGTAT	1980
GTGAGCTCCC AGTGGGAGGT GAGCCAGGTG CCCCTGGACC TGTGTGAGGT CTATGGCGGG	2040
GGCTGCCACG GTTGCCTCAT GTCCCGAGAC CCCTACTGCG GCTGGGACCA GGGCCGCTGC	2100
ATCTCCATCT ACAGCTCCGA ACGGTCAGTG CTGCAATCCA TTAATCCAGC CGAGCCACAC	2160
AAGGAGTGTC CCAACCCCAA ACCAGACAAG GCCCCACTGC AGAAGGTTTC CCTGGCCCCA	2220
AACTCTCGCT ACTACCTGAG CTGCCCCATG GAATCCCGCC ACGCCACCTA CTCATGGCGC	2280
CACAAGGAGA ACGTGGAGCA GAGCTGCGAA CCTGGTCACC AGAGCCCCAA CTGCATCCTG	2340
TTCATCGAGA ACCTCACGGC GCAGCAGTAC GGCCACTACT TCTGCGAGGC CCAGGAGGGC	2400
TCCTACTTCC GCGAGGCTCA GCACTGGCAG CTGCTGCCCC AGGACGGCAT CATGGCCGAG	2460
CACCTGCTGG GTCATGCCTG TGCCCTGGCT GCCTCCCTCT GGCTGGGGGT GCTGCCACA	2520
CTCACTCTTG GCTTGCTGGT CCACGTGAAG CTTGGGCCCC AACAAAACT CATCTCAGAA	2580
GAGGATCTGA ATAGCGCCGT CGACCATCAT CATCATCATC ATTGAGTTTA TCCAGCACAG	2640
TGGCGGCCGC TCGAGTCTAG AGGGCCCGTT TAAACCCGCT GATCAGCCTC GACTGTGCCT	2700
TCTAGTTGCC AGCCATCTGT TGTTTGCCCC TCCCCGTGC CTCCTTGAC CCTGGAAGGT	2760
GCCACTCCCA CTGTCCTTTC CTAATAAAAT GAGGAAATTG CATCGCATTG TCTGAGTAGG	2820
TGTCATTCTA TTCTGGGGGG TGGGGTGGGG CAGGACAGCA AGGGGGAGGA TTGGGAAGAC	2880
AATAGCAGGC ATGCTGGGGA TGCGGTGGGC TCTATGGCTT CTGAGGCGGA AAGAACCAGC	2940
TGGGGCTCTA GGGGGTATCC CCACGCGCCC TGTAGCGGCG CATTAAGCGC GCGGGGTGTG	3000

GTGGTTACGC GCAGCGTGAC CGCTACACTT GCCAGCGCCC TAGCGCCCCG TCCTTTCGCT	3060
TTCTTCCCTT CCTTTCCTCG CACGTTTCGCC GGCTTTCCCC GTCAAGCTCT AAATCGGGGC	3120
ATCCCTTTAG GGTTCGATT TAGTGCTTTA CGGCACCTCG ACCCCAAAAA ACTTGATTAG	3180
GGTGATGGTT CACGTAGTGG GCCATCGCCC TGATAGACGG TTTTTCGCCC TTTGACGTTG	3240
GAGTCCACGT TCTTTAATAG TGGACTCTTG TTCCAAACTG GAACAACACT CAACCCTATC	3300
TCGGTCTATT CTTTGTGATT ATAAGGGATT TTGGGGATTT CGGCCTATTG GTTAAAAAAT	3360
GAGCTGATTT AACAAAAATT TAACGCGAAT TAATTCTGTG GAATGTGTGT CAGTTAGGGT	3420
GTGGAAAGTC CCCAGGCTCC CCAGGCAGGC AGAAGTATGC AAAGCATGCA TCTCAATTAG	3480
TCAGCAACCA GGTGTGGAAA GTCCCCAGGC TCCCCAGCAG GCAGAAGTAT GCAAAGCATG	3540
CATCTCAATT AGTCAGCAAC CATAGTCCCG CCCCTAACTC CGCCCATCCC GCCCCTAACT	3600
CCGCCCAGTT CCGCCCATTC TCCGCCCCAT GGCTGACTAA TTTTTTTTAT TTATGCAGAG	3660
GCCGAGGCCG CCTCTGCCTC TGAGCTATTC CAGAAGTAGT GAGGAGGCTT TTTTGGAGGC	3720
CTAGGCTTTT GCAAAAAGCT CCCGGGAGCT TGTATATCCA TTTTCGGATC TGATCAAGAG	3780
ACAGGATGAG GATCGTTTCG CATGATTGAA CAAGATGGAT TGCACGCAGG TTCTCCGGCC	3840
GCTTGGGTGG AGAGGCTATT CGGCTATGAC TGGGCACAAC AGACAATCGG CTGCTCTGAT	3900
GCCGCCGTGT TCCGGCTGTC AGCGCAGGGG CGCCCGGTTT TTTTGTCAA GACCGACCTG	3960
TCCGGTGCCC TGAATGAACT GCAGGACGAG GCAGCGCGGC TATCGTGGCT GGCCACGACG	4020
GGCGTTCCTT GCGCAGCTGT GCTCGACGTT GTCACTGAAG CGGGAAGGGA CTGGCTGCTA	4080
TTGGGCGAAG TGCCGGGGCA GGATCTCCTG TCATCTCACC TTGCTCCTGC CGAGAAAGTA	4140
TCCATCATGG CTGATGCAAT GCGGCGGCTG CACAGCTTG ATCCGGCTAC CTGCCCATTG	4200
GACCACCAAG CGAAACATCG CATCGAGCGA GCACGTACTC GGATGGAAGC CGGTCTTGTC	4260
GATCAGGATG ATCTGGACGA AGAGCATCAG GGGCTCGCGC CAGCCGAAGT GTTCGCCAGG	4320
CTCAAGGCGC GCATGCCCCA CGGCGAGGAT CTCGTCGTGA CCCATGGCGA TGCCTGCTTG	4380
CCGAATATCA TGGTGGAAAA TGGCCGCTTT TCTGGATTCA TCGACTGTGG CCGGCTGGGT	4440
GTGGCGGACC GCTATCAGGA CATAGCGTTG GCTACCCGTG ATATTGCTGA AGAGCTTGGC	4500
GGCGAATGGG CTGACCGCTT CCTCGTGCTT TACGGTATCG CCGCTCCCGA TTCGCAGCGC	4560
ATCGCCTTCT ATCGCCTTCT TGACGAGTTC TTCTGAGCGG GACTCTGGGG TTCGAAATGA	4620
CCGACCAAGC GACGCCCAAC CTGCCATCAC GAGATTTCTG TTCCACCGCC GCCTTCTATG	4680

AAAGGTTGGG	CTTCGGAATC	GTTTTCCGGG	ACGCCGGCTG	GATGATCCTC	CAGCGCGGGG	4740
ATCTCATGCT	GGAGTTCTTC	GCCCACCCCA	ACTTGTTTAT	TGCAGCTTAT	AATGGTTACA	4800
AATAAAGCAA	TAGCATCACA	AATTTACAAA	ATAAAGCATT	TTTTTCACTG	CATTCTAGTT	4860
GTGGTTTGTC	CAAACTCATC	AATGTATCTT	ATCATGTCTG	TATACCGTCG	ACCTCTAGCT	4920
AGAGCTTGGC	GTAATCATGG	TCATAGCTGT	TTCCTGTGTG	AAATTGTTAT	CCGCTCACAA	4980
TTCCACACAA	CATACGAGCC	GGAAGCATAA	AGTGTAAGC	CTGGGGTGCC	TAATGAGTGA	5040
GCTAACTCAC	ATTAATTGCG	TTGCGCTCAC	TGCCCCGTTT	CCAGTCGGGA	AACCTGTCGT	5100
GCCAGCTGCA	TTAATGAATC	GGCCAACGCG	CGGGGAGAGG	CGGTTTGCGT	ATTGGGCGCT	5160
CTTCCGCTTC	CTCGCTCACT	GACTCGCTGC	GCTCGGTCGT	TCGGCTGCGG	CGAGCGGTAT	5220
CAGCTCACTC	AAAGGCGGTA	ATACGGTTAT	CCACAGAATC	AGGGGATAAC	GCAGGAAAGA	5280
ACATGTGAGC	AAAAGGCCAG	CAAAGGCCA	GGAACCGTAA	AAAGGCCGCG	TTGCTGGCGT	5340
TTTTCCATAG	GCTCCGCCCC	CCTGACGAGC	ATCACAAAAA	TCGACGCTCA	AGTCAGAGGT	5400
GGCGAAACCC	GACAGGACTA	TAAAGATACC	AGGCGTTTCC	CCCTGGAAGC	TCCCTCGTGC	5460
GCTCTCCTGT	TCCGACCCTG	CCGCTTACCG	GATACCTGTC	CGCCTTTCTC	CCTTCGGGAA	5520
GCGTGGCGCT	TTCTCAATGC	TCACGCTGTA	GGTATCTCAG	TTCGGTGTAG	GTCGTTGCGT	5580
CCAAGCTGGG	CTGTGTGCAC	GAACCCCCCG	TTCAGCCCGA	CCGCTGCGCC	TTATCCGGTA	5640
ACTATCGTCT	TGAGTCCAAC	CCGGTAAGAC	ACGACTTATC	GCCACTGGCA	GCAGCCACTG	5700
GTAACAGGAT	TAGCAGAGCG	AGGTATGTAG	GCGGTGCTAC	AGAGTTCTTG	AAGTGGTGGC	5760
CTAACTACGG	CTACACTAGA	AGGACAGTAT	TTGGTATCTG	CGCTCTGCTG	AAGCCAGTTA	5820
CCTTCGGAAA	AAGAGTTGGT	AGCTCTTGAT	CCGGCAAACA	AACCACCGCT	GGTAGCGGTG	5880
GTTTTTTTGT	TTGCAAGCAG	CAGATTACGC	GCAGAAAAAA	AGGATCTCAA	GAAGATCCTT	5940
TGATCTTTTC	TACGGGGTCT	GACGCTCAGT	GGAACGAAAA	CTCACGTTAA	GGGATTTTGG	6000
TCATGAGATT	ATCAAAAAGG	ATCTTCACCT	AGATCCTTTT	AAATTAAAAA	TGAAGTTTTA	6060
AATCAATCTA	AAGTATATAT	GAGTAAACTT	GGTCTGACAG	TTACCAATGC	TTAATCAGTG	6120
AGGCACCTAT	CTCAGCGATC	TGTCTATTTT	GTTTCATCCAT	AGTTGCCTGA	CTCCCCGTG	6180
TGTAGATAAC	TACGATACGG	GAGGGCTTAC	CATCTGGCCC	CAGTGCTGCA	ATGATACCGC	6240
GAGACCCACG	CTCACC GGCT	CCAGATTTAT	CAGCAATAAA	CCAGCCAGCC	GGAAGGGCCG	6300
AGCGCAGAAG	TGGTCCTGCA	ACTTTATCCG	CCTCCATCCA	GTCTATTAAT	TGTTGCCGGG	6360
AAGCTAGAGT	AAGTAGTTCG	CCAGTTAATA	GTTTGCGCAA	CGTTGTTGCC	ATTGCTACAG	6420

GCATCGTGGT GTCACGCTCG TCGTTTGGTA TGGCTTCATT CAGCTCCGGT TCCCAACGAT	6480
CAAGGCGAGT TACATGATCC CCCATGTTGT GCAAAAAAGC GGTTAGCTCC TTCGGTCCTC	6540
CGATCGTTGT CAGAAGTAAG TTGGCCGCAG TGTTATCACT CATGGTTATG GCAGCACTGC	6600
ATAATTCTCT TACTGTCATG CCATCCGTAA GATGCTTTTC TGTGACTGGT GAGTACTCAA	6660
CCAAGTCATT CTGAGAATAG TGTATGCGGC GACCGAGTTG CTCTTGCCCG GCGTCAATAC	6720
GGGATAATAC CGCGCCACAT AGCAGAACTT TAAAGTGCT CATCATTGGA AAACGTTCTT	6780
CGGGGCGAAA ACTCTCAAGG ATCTTACCGC TGTTGAGATC CAGTTCGATG TAACCCACTC	6840
GTGCACCCAA CTGATCTTCA GCATCTTTTA CTTTCACCAG CGTTTCTGGG TGAGCAAAAA	6900
CAGGAAGGCA AAATGCCGCA AAAAAGGGAA TAAGGGCGAC ACGGAAATGT TGAATACTCA	6960
TACTCTTCCT TTTTCAATAT TATTGAAGCA TTTATCAGGG TTATTGTCTC ATGAGCGGAT	7020
ACATATTTGA ATGTATTTAG AAAAATAAAC AAATAGGGGT TCCGCGCACA TTTCCCCGAA	7080
AAGTGCCACC TGACGTCGAC GGATCGGG	7108

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4019 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CTCGAGAAAT CATAAAAAAT TTATTTGCTT TGTGAGCGGA TAACAATTAT AATAGATTCA	60
ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGAGA	120
GGATCGCATC ACCATCACCA TCACGGATCC CTGGTTCTGT TTGAAGGGGA CGAGGTGTAT	180
TCCACCATCC GGAAGCAGGA ATACAATGGG AAGATCCCTC GGTTCGCGCG CATCCGGGGC	240
GAGAGTGAGC TGTACACCAG TGATACTGTC ATGCAGAACC CACAGTTCAT CAAAGCCACC	300
ATCGTGCACC AAGACCAGGC TTACGATGAC AAGATCTACT ACTTCTTCCG AGAGGACAAT	360
CCTGACAAGA ATCCTGAGGC TCCTCTCAAT GTGTCCCGTG TGGCCAGTT GTGCAGGGGG	420
GACCAGGGTG GGGAAAGTTC ACTGTCAGTC TCCAAGTGGA ACACTTTTCT GAAAGCCATG	480
CTGGTATGCA GTGATGCTGC CACCAACAAG AACTTCAACA GGCTGCAAGA CGTCTTCCTG	540

CTCCCTGACC CCAGCGGCCA GTGGAGGGAC ACCAGGGTCT ATGGTGTTTT CTCCAACCCC	600
TGGAAC TACT CAGCCGTCTG TGTGTATTCC CTCGGTGACA TTGACAAGGT CTTCCGTACC	660
TCCTCACTCA AGGGCTACCA CTCAAGCCTT CCCAACCCGC GGCCTGGCAA GTGCCTCCCA	720
GACCAGCAGC CGATACCCAC AGAAAGCTTA ATTAGCTGAG CTTGGACTCC TGTTGATAGA	780
TCCAGTAATG ACCTCAGAAC TCCATCTGGA TTTGTTT CAGA ACGCTCGGTT GCCGCCGGG	840
GTTTTTTTATT GGTGAGAATC CAAGCTAGCT TGGCGAGATT TTCAGGAGCT AAGGAAGCTA	900
AAATGGAGAA AAAAATCACT GGATATACCA CCGTTGATAT ATCCCAATGG CATCGTAAAG	960
AACATTTTGA GGCATTTT CAG TCAGTTGCTC AATGTACCTA TAACCAGACC GTTCAGCTGG	1020
ATATTACGGC CTTTTTAAAG ACCGTAAAGA AAAATAAGCA CAAGTTTAT CCGGCCTTTA	1080
TTCACATTCT TGCCCGCCTG ATGAATGCTC ATCCGGAATT TCGTATGGCA ATGAAAGACG	1140
GTGAGCTGGT GATATGGGAT AGTGTTTACC CTTGTTACAC CGTTTTCCAT GAGCAAAC T	1200
AAACGTTTTT C ATCGCTCTGG AGTGAATACC ACGACGATTT CCGGCAGTTT CTACACATAT	1260
ATTTCGAAGA TGTGGCGTGT TACGGTGAAA ACCTGGCCTA TTTCCCTAAA GGGTTTATTG	1320
AGAATATGTT TTTCGTCTCA GCCAATCCCT GGGTGAGTTT CACCAGTTTT GATTTAAACG	1380
TGGCCAATAT GGACAAC TTCGCCCCCG TTTTCACCAT GGGCAAATAT TATACGCAAG	1440
GCGACAAGGT GCTGATGCCG CTGGCGATTC AGGTTCATCA TGCCGTCTGT GATGGCTTCC	1500
ATGTCGGCAG AATGCTTAAT GAATTACAAC AGTACTGCGA TGAGTGGCAG GGCGGGGCGT	1560
AATTTTTTTTA AGGCAGTTAT TGGTGCCCTT AAACGCCTGG GGTAAATGACT CTCTAGCTTG	1620
AGGCATCAAA TAAAACGAAA GGCTCAGTCG AAAGACTGGG CCTTTCGTTT TATCTGTTGT	1680
TTGTCGGTGA ACGCTCTCCT GAGTAGGACA AATCCGCCGC TCTAGAGCTG CCTCGCGCGT	1740
TTCGGTGATG ACGGTGAAAA CCTCTGACAC ATGCAGCTCC CGGAGACGGT CACAGCTTGT	1800
CTGTAAGCGG ATGCCGGGAG CAGACAAGCC CGTCAGGGCG CGTCAGCGGG TGTGCGGGG	1860
TGTCGGGGCG CAGCCATGAC CCAGTCACGT AGCGATAGCG GAGTGTATAC TGGCTTAACT	1920
ATGCGGCATC AGAGCAGATT GTACTGAGAG TGCACCATAT GCGGTGTGAA ATACCGCACA	1980
GATGCGTAAG GAGAAAATAC CGCATCAGGC GCTCTTCCGC TTCCTCGCTC ACTGACTCGC	2040
TGCGCTCGGT CTGTCGGCTG CGGCGAGCGG TATCAGCTCA CTCAAAGGCG GTAATACGGT	2100
TATCCACAGA ATCAGGGGAT AACGCAGGAA AGAACATGTG AGCAAAAGGC CAGCAAAAGG	2160
CCAGGAACCG TAAAAAGGCC GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCCTGACG	2220

AGCATCACAA AAATCGACGC TCAAGTCAGA GGTGGCGAAA CCCGACAGGA CTATAAAGAT	2280
ACCAGGCGTT TCCCCCTGGA AGCTCCCTCG TGCGCTCTCC TGTTCCGACC CTGCCGCTTA	2340
CCGGATACCT GTCCGCCTTT CTCCCTTCGG GAAGCGTGGC GCTTTCTCAA TGCTCACGCT	2400
GTAGGTATCT CAGTTCGGTG TAGGTCGTTT GCTCCAAGCT GGGCTGTGTG CACGAACCCC	2460
CCGTTTCAGCC CGACCGCTGC GCCTTATCCG GTAACATATCG TCTTGAGTCC AACCCGGTAA	2520
GACACGACTT ATCGCCACTG GCAGCAGCCA CTGGTAACAG GATTAGCAGA GCGAGGTATG	2580
TAGGCGGTGC TACAGAGTTC TTGAAGTGGT GGCCTAACTA CGGCTACACT AGAAGGACAG	2640
TATTTGGTAT CTGCGCTCTG CTGAAGCCAG TTACCTTCGG AAAAAGAGTT GGTAGCTCTT	2700
GATCCGGCAA ACAAACCACC GCTGGTAGCG GTGGTTTTTT TGTTTGCAAG CAGCAGATTA	2760
CGCGCAGAAA AAAAGGATCT CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC	2820
AGTGGAACGA AAACCTACGT TAAGGGATTT TGGTCATGAG ATTATCAAAA AGGATCTTCA	2880
CCTAGATCCT TTAAATTAA AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA	2940
CTTGGTCTGA CAGTTACCAA TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT	3000
TTCGTTTCATC CATAGCTGCC TGAATCCCCG TCGTGTAGAT AACTACGATA CGGGAGGGCT	3060
TACCATCTGG CCCCAGTGCT GCAATGATAC CGCGAGACCC ACCTCACCAG GCTCCAGATT	3120
TATCAGCAAT AAACCAGCCA GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT GCAACTTTAT	3180
CCGCCTCCAT CCAGTCTATT AATTGTTGCC GGAAGCTAG AGTAAGTAGT TCGCCAGTTA	3240
ATAGTTTGCG CAACGTTGTT GCCATTGCTA CAGGCATCGT GGTGTCACGC TCGTCGTTTG	3300
GTATGGCTTC ATTCAGCTCC GGTTCCTAAC GATCAAGGCG AGTTACATGA TCCCCATGT	3360
TGTGCAAAAA AGCGGTTAGC TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG	3420
CAGTGTTATC ACTCATGGTT ATGGCAGCAC TGCATAATC TCTTACTGTC ATGCCATCCG	3480
TAAGATGCTT TTCTGTGACT GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGATGTC	3540
GGCGACCGAG TTGCTCTTGC CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA	3600
CTTTAAAGT GCTCATCATT GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC	3660
CGCTGTTGAG ATCCAGTTCG ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT	3720
TTACTTTCAC CAGCGTTTCT GGGTGAGCAA AACAGGAAG GCAAAATGCC GCAAAAAGG	3780
GAATAAGGGC GACACGGAAA TGTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA	3840
GCATTTATCA GGGTTATTGT CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA	3900
AACAAATAGG GGTTCGCGC ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA	3960

TTATTATCAT GACATTAACC TATAAAAATA GCGTATCAC GAGGCCCTTT CGTCTTCAC 4019

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3999 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

CTCGAGAAAT CATAAAAAAT TTATTTGCTT TGTGAGCGGA TAACAATTAT AATAGATTCA	60
ATTGTGAGCG GATAACAATT TCACACAGAA TTCATTAAAG AGGAGAAATT AACTATGAGA	120
GGATCGCATC ACCATCACCA TCACACGGAT CCGCATGCGA GCTCCCAGTG GGAGGTGAGC	180
CAGGTGCCCC TGGACCTGTG TGAGGTCTAT GCGGGGGGCT GCCACGGTTG CCTCATGTCC	240
CGAGACCCCT ACTGCGGCTG GGACCAGGGC CGCTGCATCT CCATCTACAG CTCCGAACGG	300
TCAGTGCTGC AATCCATTAA TCCAGCCGAG CCACACAAGG AGTGTCCCAA CCCCAAACCA	360
GACAAGGCCC CACTGCAGAA GGTTTCCCTG GCCCCAACT CTCGCTACTA CCTGAGCTGC	420
CCCATGGAAT CCCGCCACGC CACCTACTCA TGGCGCCACA AGGAGAACGT GGAGCAGAGC	480
TGCGAACCTG GTCACCAGAG CCCCAACTGC ATCCTGTTCA TCGAGAACCT CACGGCGCAG	540
CAGTACGGCC ACTACTTCTG CGAGGCCAG GAGGGCTCCT ACTTCCGCGA GGCTCAGCAC	600
TGGCAGCTGC TGCCCGAGGA CGGCATCATG GCCGAGCACC TGCTGGGTCA TGCCTGTGCC	660
CTGGCTGCCT CCCTCTGGCT GGGGGTGCTG CCCCACTCA CTCTTGGCTT GCTGGTCCAC	720
GTGAAGCTTA ATTAGCTGAG CTTGGACTCC TGTTGATAGA TCCAGTAATG ACCTCAGAAC	780
TCCATCTGGA TTTGTTTCTG ACGCTCGGTT GCCGCCGGGC GTTTTTTATT GGTGAGAATC	840
CAAGCTAGCT TGGCGAGATT TTCAGGAGCT AAGGAAGCTA AAATGGAGAA AAAAATCACT	900
GGATATACCA CCGTTGATAT ATCCCAATGG CATCGTAAAG AACATTTTGA GGCATTTCAG	960
TCAGTTGCTC AATGTACCTA TAACCAGACC GTTCAGCTGG ATATTACGGC CTTTTTAAAG	1020
ACCGTAAAGA AAAATAAGCA CAAGTTTTAT CCGGCCTTTA TTCACATTCT TGCCCGCCTG	1080
ATGAATGCTC ATCCGGAATT TCGTATGGCA ATGAAAGACG GTGAGCTGGT GATATGGGAT	1140
AGTGTTACAC CTTGTTACAC CGTTTTCCAT GAGCAAACCTG AAACGTTTTT ATCGCTCTGG	1200

AGTGAATACC ACGACGATTT CCGGCAGTTT CTACACATAT ATTGCAAGA TGTGGCGTGT	1260
TACGGTGAAA ACCTGGCCTA TTTCCCTAAA GGGTTTATTG AGAATATGTT TTTCGTCTCA	1320
GCCAATCCCT GGGTGAGTTT CACCAGTTTT GATTAAACG TGGCCAATAT GGACAACTTC	1380
TTGCCCCCG TTTTCACCAT GGGCAAATAT TATACGCAAG GCGACAAGGT GCTGATGCCG	1440
CTGGCGATT C AGGTTTCATCA TGCCGTCTGT GATGGCTTCC ATGTCGGCAG AATGCTTAAT	1500
GAATTACAAC AGTACTGCGA TGAGTGGCAG GCGGGGGCGT AATTTTTTTTA AGGCAGTTAT	1560
TGGTGCCCTT AAACGCCTGG GGTAATGACT CTCTAGCTTG AGGCATCAAA TAAAACGAAA	1620
GGCTCAGTCG AAAGACTGGG CCTTTCGTTT TATCTGTTGT TTGTCGGTGA ACGCTCTCCT	1680
GAGTAGGACA AATCCGCCGC TCTAGAGCTG CCTCGCGCGT TTCGGTGATG ACGGTGAAAA	1740
CCTCTGACAC ATGCAGCTCC CGGAGACGGT CACAGCTTGT CTGTAAGCGG ATGCCGGGAG	1800
CAGACAAGCC CGTCAGGGCG CGTCAGCGGG TGTTGGCGGG TGTCGGGGCG CAGCCATGAC	1860
CCAGTCACGT AGCGATAGCG GAGTGTATAC TGGCTTAAC TATGCGGCATC AGAGCAGATT	1920
GTACTGAGAG TGCACCATAT GCGGTGTGAA ATACCGCACA GATGCGTAAG GAGAAAATAC	1980
CGCATCAGGC GCTCTTCCGC TTCCTCGCTC ACTGACTCGC TCGCTCGGT CTGTCGGCTG	2040
CGGCGAGCGG TATCAGCTCA CTCAAAGGCG GTAATACGGT TATCCACAGA ATCAGGGGAT	2100
AACGCAGGAA AGAACATGTG AGCAAAAGGC CAGCAAAAGG CCAGGAACCG TAAAAAGGCC	2160
GCGTTGCTGG CGTTTTTCCA TAGGCTCCGC CCCCTGACG AGCATCACAA AAATCGACGC	2220
TCAAGTCAGA GGTGGCGAAA CCCGACAGGA CTATAAAGAT ACCAGGCGTT TCCCCCTGGA	2280
AGCTCCCTCG TGCCTCTCC TGTTCCGACC CTGCCGCTTA CCGGATACCT GTCCGCCTTT	2340
CTCCCTTCGG GAAGCGTGGC GCTTTCTCAA TGCTCACGCT GTAGGTATCT CAGTTCGGTG	2400
TAGGTCGTT C GCTCCAAGCT GGGCTGTGTG CACGAACCCC CCGTTCAGCC CGACCGCTGC	2460
GCCTTATCCG GTAACATCG TCTTGAGTCC AACCCGGTAA GACACGACTT ATCGCCACTG	2520
GCAGCAGCCA CTGGTAACAG GATTAGCAGA GCGAGGTATG TAGGCGGTGC TACAGAGTTC	2580
TTGAAGTGGT GGCCTAACTA CGGCTACACT AGAAGGACAG TATTTGGTAT CTGCGCTCTG	2640
CTGAAGCCAG TTACCTTCGG AAAAAGAGTT GGTAAGCTCTT GATCCGGCAA ACAAACCACC	2700
GCTGGTAGCG GTGGTTTTTT TGTTTGCAAG CAGCAGATTA CGCGCAGAAA AAAAGGATCT	2760
CAAGAAGATC CTTTGATCTT TTCTACGGGG TCTGACGCTC AGTGGAACGA AAACTCACGT	2820
TAAGGGATTT TGGTCATGAG ATTATCAAAA AGGATCTTCA CCTAGATCCT TTAAATTAA	2880

AAATGAAGTT TTAAATCAAT CTAAAGTATA TATGAGTAAA CTTGGTCTGA CAGTTACCAA	2940
TGCTTAATCA GTGAGGCACC TATCTCAGCG ATCTGTCTAT TTCGTTTCATC CATAGCTGCC	3000
TGACTCCCCG TCGTGTAGAT AACTACGATA CGGGAGGGCT TACCATCTGG CCCCAGTGCT	3060
GCAATGATAC CGCGAGACCC ACGCTCACCG GCTCCAGATT TATCAGCAAT AAACCAGCCA	3120
GCCGGAAGGG CCGAGCGCAG AAGTGGTCCT GCAACTTTAT CCGCCTCCAT CCAGTCTATT	3180
AATTGTTGCC GGGAAGCTAG AGTAAGTAGT TCGCCAGTTA ATAGTTTGCG CAACGTTGTT	3240
GCCATTGCTA CAGGCATCGT GGTGTCACGC TCGTCGTTTG GTATGGCTTC ATTCAGCTCC	3300
GGTTCCCAAC GATCAAGGCG AGTTACATGA TCCCCATGT TGTGCAAAAA AGCGGTTAGC	3360
TCCTTCGGTC CTCCGATCGT TGTCAGAAGT AAGTTGGCCG CAGTGTATC ACTCATGGTT	3420
ATGGCAGCAC TGCATAATTC TCTTACTGTC ATGCCATCCG TAAGATGCTT TTCTGTGACT	3480
GGTGAGTACT CAACCAAGTC ATTCTGAGAA TAGTGTATGC GGCGACCGAG TTGCTCTTGC	3540
CCGGCGTCAA TACGGGATAA TACCGCGCCA CATAGCAGAA CTTTAAAAGT GCTCATCATT	3600
GGAAAACGTT CTTCGGGGCG AAAACTCTCA AGGATCTTAC CGCTGTTGAG ATCCAGTTCG	3660
ATGTAACCCA CTCGTGCACC CAACTGATCT TCAGCATCTT TTACTTTCAC CAGCGTTTCT	3720
GGGTGAGCAA AAACAGGAAG GCAAAATGCC GCAAAAAAGG GAATAAGGGC GACACGGAAA	3780
TGTTGAATAC TCATACTCTT CCTTTTTCAA TATTATTGAA GCATTTATCA GGGTTATTGT	3840
CTCATGAGCG GATACATATT TGAATGTATT TAGAAAAATA AACAAATAGG GGTCCGCGC	3900
ACATTTCCCC GAAAAGTGCC ACCTGACGTC TAAGAAACCA TTATTATCAT GACATTAACC	3960
TATAAAAATA GCGGTATCAC GAGGCCCTTT CGTCTTCAC	3999

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8888 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

GAGCCGCACA CGGTGCTTTT CCACGAGCCA GGCAGCTCCT CTGTGTGGGT GGGAGGACGT	60
GGCAAGGTCT ACCTCTTTGA CTTCCCCGAG GGCAAGAACG CATCTGTGCG CACGGTGAGC	120

CTCTCTCTTC CCCAACACCC CCCCCTACCC TCTTATCTCC CCTCTGGCCC TGCCAAGGGT	180
CCTCAGGGAA TCCGAGGGAG CTGGCTTCTC TTCCTAAACT GCCCCACCT CCGTATCCTA	240
TAAATGGCTC CTGGGGGAGG CTCCCTAAAG GTAGTCCAGA TTGGAGTGGG GAGCTGGGGC	300
GGTGTGGAGA AAAACAGGAG CTAATGGGCC TGGCCAGCTG GGCAGCGCTG CTGCGGAAAG	360
CCCAGGCTGG AAGCTGGGCC CCAGAGCCCA TGCCTGGTCT TCTGAACCCT CTGGGCCTCA	420
GCTCTGGATA TGAGACCCTG TTTGACCTCA GGTAGATCAC TCACCCTCTC AGAGCCCCAG	480
TTGCTCATCT GTCAGATGAG AATAATGGTT GCTTCCTTTG GGGCTTATCC TGAGGCTGTG	540
TGGAAAGCAT TTCAGGGGTA CCTCACCCCT GGCAGATTGA ACTAATGCTT CTCCCCTTCC	600
CCAGGTGAAT ATCGGCTCCA CAAAGGGGTC CTGTCTGGAT AAGCGGGTGA GCGGGGGAGG	660
GATCTGGAGG GGTCTGAGCC ACTTGGTAAA GGGAGAGGAG ACCCTGAGGG TCTAAGGAAG	720
GAAGCATGGC CCTGCCCCAC GAGTCCCAGA CTGATGGGGA GACGTGGTCC TCTGTGCTTA	780
GGGGATGGCG TCAGCTGCAC AACTCTGGG CTGTCCCGGG AGGCTGTCAC CTATGCTAAG	840
CCCTTCTGAC ACCTTCTTCC CTGATCCTGG GGGTCCTAGT GCTAGGCTTG CCAGGGCCTT	900
CCAGCAACCA ATTTCTCTCC TCCCTTCTCT CTTCCCCGGG CAGGACTGCG AGAACTACAT	960
CACCTCTCTG GAGAGGCGGA GTGAGGGGCT GCTGGCCTGT GGCACCAACG CCCGGCACCC	1020
CAGCTGCTGG AACCTGGTGA GAAGGCTGCT CCCCATGTGC CTGATCAGCT CACCTTCTAC	1080
TGCGTGGGCT TCTGCCCCTC ATGGTGGGAA GGAGATGGCG AGACTCCAAT GCTGGCCTTG	1140
CCCTGGGAGG ATGGGGCTCC TGGCCGAGAA ACTGGCCGTC ATGGGAGGCA GTGGCTGTGG	1200
GATTATGTGG CCATCCAACC CTCTGGATCT CCCACAGGTG AATGGCACTG TGGTGCCACT	1260
TGGCGAGATG AGAGGCTACG CCCCCTTCAG CCCGGACGAG AACTCCCTGG TTCTGTTTGA	1320
AGGTTGGGGC ATGCTTCGGA ACTGGGCTGG GAGCAGGATG GTCAGCTCTT TGTCCAGTGT	1380
CCGGAGGAGG GACTTCCAGG AGCTGCCTGC CCTTACTCAT TTCTCCCTCC CACTGACCCC	1440
AGGGGACGAG GTGTATTCCA CCATCCGGAA GCAGGAATAC AATGGGAAGA TCCCTCGGTT	1500
CCGCCGCATC CGGGGCGAGA GTGAGCTGTA CACCAGTGAT ACTGTCATGC AGAGTGAGTC	1560
AGGCTCCGGC TGGGCTGAGG GTGGGCAAGG GGGTGTGAGC ACTTAAGGTG GCAGATGGGA	1620
TCCTGATGTT TCTGGGAGGG CTCCCTGAGG GCCGCTGGGG CCATGCAGGA AAGCAGGACC	1680
TTGGTATAGG CCTGAGAAGT TAGGGTTGGC TGGGAGCAGA GGAACAGACA AGGTATAGCA	1740
GTGGGATGGG CCCAGCCCTC TTCAGGAACA CAAACAGAGG GAGCCCCAGA CCCAGTGCAG	1800
GGTCCCCAGG AGCCAAAGTT TATCCTCTGC TGAGTTCACG TGGAGGCAGC CCCCCAACTC	1860

CCTCCTCATC	AGGGCTCTGC	CAATTGAGCA	GAAGTGACAT	AGGGGCCCCC	AGGGACCTTC	1920
CCCCACTCCC	CAGGCATGAA	GTCATTGCTC	CTGGGCCGAT	GACATCTTTG	TAGGAAGAGG	1980
GCAAAACAGG	TGTGGGGTGG	AGGTGCAGGG	TCTAGGGCCC	CTCGGGGAGT	TGGACCTGAT	2040
GTTATGAGTC	CTATTCCAGA	TCTGATTTGC	CATGGTTTGT	GCAGACCCGA	AGGAGGGAGG	2100
AGAGTGTGCA	GGGTTGGAAT	GGTCTCCCCG	GCAAGCTTCC	CAGCCTTACG	CCCATTGCTC	2160
TCTGTGCCCT	GGCAGACCCA	CAGTTCATCA	AAGCCACCAT	CGTGCACCAA	GACCAGGCTT	2220
ACGATGACAA	GATCTACTAC	TTCTTCCGAG	AGGACAATCC	TGACAAGAAT	CCTGAGGCTC	2280
CTCTCAATGT	GTCCCGTGTG	GCCCAGTTGT	GCAGGGTGAA	CACGGGCGTG	AGGGCTGCTG	2340
GCTACGTGTC	TGTGCATGAA	TAGGCCTGAG	TGAGGGTGAG	TTCTGTGTGT	CCGTGTGCAT	2400
GTAGAAGTTG	TGTGGATGTA	TGAGTGGGTC	TGTGTCAGGG	ACTGTGGGAG	CAGCTGTGTG	2460
TGCATGGAGC	ATCATGTGTC	TGTGTGTGGG	TAAAGGTGGC	TGAGCTCCTG	TGCACGTATG	2520
ATGGCGTGTG	AGCGTGTGTA	TGATGGGGTG	TGTGTGTGTG	TGTGTGTGTG	TGTTTTGCCT	2580
GTGTGAATGT	GCTGTGCCAC	GTATGTGGGT	GCGTGAGTCA	GTAAATGTGT	GTCTGAGTCC	2640
GTCTGCTCTG	TGGGGACCTG	GCACTCTCAC	CTGCCCTGAC	CCTGGGCACT	GCTGGCCCTG	2700
GGCTCTGGAT	CAGCCAGGCC	TGCTTGCAGG	AGTCTCATCT	GGAGACCTGC	CCTGAGTCCT	2760
GGGGCACCCC	CGGCAGGTCC	TGGCCCCCTG	CAGCCTGCCT	TCCTCCTCTG	GGCCCAGGTG	2820
TTGATATTGC	TGGCAGTGGT	TTCTTGGGGT	GTGTGGGGAA	GCCCCGGCAG	GTGCTGAGGG	2880
GCCTCTTCTC	CCCTCTACCC	TTCCAGGGGG	ACCAGGGTGG	GGAAAGTTCA	CTGTCAGTCT	2940
CCAAGTGGA	CACTTTTCTG	AAAGCCATGC	TGGTATGCAG	TGATGCTGCC	ACCAACAAGA	3000
ACTTCAACAG	GCTGCAAGAC	GTCTTCCTGC	TCCCTGACCC	CAGCGGCCAG	TGGAGGGACA	3060
CCAGGGTCTA	TGGTGTTTTC	TCCAACCCCT	GGTGAGTGGC	CCTTGTCCCTG	GGGCCGGGGC	3120
TGGCATTGGT	TCAGTGTCCA	GTAGGGACAG	GAGGCCTTGG	GCCCTGCTGA	GGGCCTCCCT	3180
GGTGTGGCAG	GAGCAGGGGC	TGCAGGCTCA	AGAGGCTGGG	CTGTTGCTGG	GTGTGGGGTG	3240
GGGGGACAGC	CAGTGCGATG	TATGTACTGT	TGTGTGAGTG	AGTCTGCACT	CATGGGTGTG	3300
TGTGCATGCC	CTATATGCAC	ACTCATGACT	GCACTTGTGC	CTGTGTGTCC	CACCACCTGC	3360
TTGTGCCGAG	AGTGGACACT	GGGCCCAGGA	GGAAGCTGCT	GAAGCATCTC	TCGGGGAGCT	3420
GGGTGCTATT	ACACCTGCTC	AGGCACTGCC	TGAGCCCGAT	AATTCACACT	TCTTAATCAC	3480
TCTCATTGAT	TGAACACACG	GCAGGCGGAA	GTGTTGGGTG	TGTGTGGGGA	GAGTTAGGGA	3540

TAGAGTGGAG GAAGCCAAGA CCCTGCTCTG TGGCTCCTGG GTGAGTGGGT CCCCCAGGCT	3600
GGGAAGGGGT TGGGGGTCTG GCCTCCTGGG GCATCAGCAC CCCACAGCCT GTGCCCAGGG	3660
AGGGCTAGAG AACTGCTCAG CCTATGATGG GGTTCCTCCT GCCTTGGGGT TGGGTAGAGC	3720
AGATGGCCTC TAGACTCAGT GATTCTGTAA CAGGATACAA GTTTGTGGTT TTAAATTGCA	3780
GCACAAAGAA ATTAGGCTGA ACTCCTCTCC TTCCTCCTCT CCATCCCTCC CCATTTTCAG	3840
TGGTGGTTGG CAACTCAGTG CCAGGCACAA GGCTGGCCTG GGTGAGTGGA GGTGGATGGG	3900
TGGGTTCCTG GCCCCCATT GAGCTGGTCT CCATGTCACT GCAGGAAC TA CTCAGCCGTC	3960
TGTGTGTATT CCCTCGGTGA CATTGACAAG GTCTTCCGTA CCTCCTCACT CAAGGGCTAC	4020
CACTCAAGCC TTCCCAACCC GCGGCCTGGC AAGGTGAGCG TGACACCAGC CGTGGCCCAG	4080
GCCCAGCCCT CTTTCTGCCT CACCTCCCAC CACCCCACTG ACCTGGGCCT GCTCTCCTTG	4140
CCCAGTGCCT CCCAGACCAG CAGCCGATAC CCACAGAGAC CTTCCAGGTG GCTGACCGTC	4200
ACCCAGAGGT GCGCAGAGG GTGGAGCCCA TGGGGCCTCT GAAGACGCCA TTGTTCCTACT	4260
CTAAATACCA CTACCAGAAA GTGGCCGTCC ACCGCATGCA AGCCAGCCAC GGGGAGACCT	4320
TTCATGTGCT TTACCTAACT ACAGGTGAGA GGCTACCCCG GGACCCCTCAG TTTGCTTTGT	4380
AAAAACGGGC ATGAAAGGTG TAAGGAATAA TGAGTTAAC ATCTGGTTGG ATCTTTACAT	4440
GTGGAAGGAA TAATTGAGTG ACTGGAGTTG TCAGGGGTTA ATGTGTGTGG GTGTGGAAGA	4500
GCCAGGCAGG GAGAGCTTCC TGGAGGAGGT AGGGGCAAGA GGGAAAGGGG GATGGGAGAA	4560
AAGCAAGCAC TGGGATTTGG AGGCGGAAAT CTGGAGAGTC TGAGCAAAGC CAGGTGCACC	4620
TTTGGTCCAG ATGTCTGACT CAGGGAAGAA GATGGTAGGA AGAGACGTGG CAAATGAGGA	4680
GGAGGGGCCT GAACCACAGG GATACTGGCC TCTGCCAGGC AGAATGAGGG AGTCAGGCCC	4740
TGCGCCTGTC TTTGGGATTG TGCAGGTGAG AAGAAACATT TGAGGAGTTG ATGGGGCACA	4800
AATTAGGTAT GGGGAAGGAG TTCCAGGGGG CAGAACCTTT GCCATCTCAC AGAGGACAGG	4860
GGCAGCTTCT CTTCTTCCCT GGAGTAGGCC CTGCTGGGGG AAGCTGGGTG GAATGCCGTG	4920
GGAGATGCTC CTGCTTTCTG GAAAGCCACA GGACACGGAG GAGCCAGTCC TGAGTTGGGT	4980
TTGTGCGAGC TTCCCATGCC AGCTGCCTTC CTTGAGACTG GAAAGGGCCT CTAGCACCCC	5040
TGGGGCCATT CAATTCAGGC CCAGGCGCCC AACCTCAGTT GTTCACATTC CCCATGTGAT	5100
CTCCTGTTGC TGCTTCACCT TGGGACTGTC TCGGCTTTGG TGACCTTGTA GGAAACTGGA	5160
ACCCAGCAC CATTGTTTGG CTCCTGGAAG CCTTGGGGAG AGGAATTTCC CACAGGGCAG	5220
GGCCTGGGTC CTGATTCCCT GCCTCTTTAC TCCCTATTCA TCCCGGCTAC ACCCTTGGGC	5280

CCCCATCCTT	GCTTGGCTCC	AGTACTGGCT	GGCACAGCTG	TTGTGGTCAT	CCAGGGATGG	5340
CAGGGCACTG	GGGAACAGAA	GAGAGAGGTC	ACACAGTGCG	GAACTGGGAG	CAGGAGCTAG	5400
GACAAGGAAG	GCTGGACTTG	GGCCATGGAT	TCCCTTCCTG	CAGACTTGGG	AAGTGAGCAG	5460
ACTTGAGTGA	TTAGAGAAGG	TGTCTTCGTT	CTAAGGGCAG	TGGAGGAGGC	ACCATTTTGG	5520
AGCCTGCATC	ATTCGTATTT	GGGCTAGATT	GAAAAATAGA	GCTTTCCTAAG	TCCTCTGCAG	5580
AGAATGGGAG	GCTCTCACAA	CTGGGAGAAG	TATTGGCTCT	TTTCCTGAGA	ATTTTGCCAA	5640
GGGTATGCTG	TTACTGGGGC	TGGTTTGGA	GGAGTATAGG	GCATTATGTC	TGTGAAGGCA	5700
GTGGCTGGGG	TGGGGCCTTA	TCAGGCCCAA	GGAGCATCTG	GCCACATCTC	AGAGTCCACA	5760
GATGAGGATC	ACGGATGTGT	AGAGGAAACA	TCCTAGGCAG	GCAATCATCT	GA CTGCTTTT	5820
TTGGGGCAGG	TGATGCCCTG	GGAAATTGGG	AGGGAGGGAG	AGAGGGAGGT	AGGCTATTCT	5880
AGAAACTGGG	AGAGCAGGTG	AGGTAGGATT	GGGAGGACCA	GGGGTCAGGG	TCCCCATTGG	5940
TCCCTAATTG	AGAACGGAGA	GAGCATTGGT	CTAGGAGGCA	GGCAGCTCGG	TTATAAGACC	6000
TTGGGAACTC	TTGATTTAGA	ATCCAAGATC	CTTTT TAGAT	CTAGGATTTT	ATAAAATTAA	6060
GATATCCCCT	AAGATCAAAT	GCAACGTGGA	GTCCTGAATT	GGATCCTAGA	ACAGAAGAAG	6120
GACATTTGTG	GAAAAACTAG	TGAAATCCAA	ATAAAGTCTG	TAGTTTTGTT	AATAGTAATG	6180
CACCAATGTC	AGTTGCCTAG	TTGTGACAAA	TATACCGTGG	TTATGTAAGA	TGGTAACATT	6240
AGGGGGAACT	GGAGAAGGGT	AGATTGGAGC	TCTCTGTACT	ATCTTTGCAA	CTTTTCTGGG	6300
AATCTAAAT	TACTCCAAAA	TAAAAAAAAA	ATGTATTTAA	AGTAAATATA	TTCCCTAAGA	6360
GTCCAGGAGG	CAGGGGAGTT	GTAGAAGCAG	CTGAGTGGTT	GGGTTC TGAC	AGATTTGGTT	6420
CCAACTCGGT	CTCTGCTGCT	CACCAGCTGT	GTGACCTTGA	GCAAGTGGCT	TAGCCTTTCT	6480
GAGCCTGATT	TCCTTATCTG	TGGAGTGGGG	AAGATGACAG	CCACCTCGCA	GGGCTGTGGA	6540
GGGTAAACG	AGGTGATGCA	TGGACAGCAG	CCGCACTGAC	CTTGCTGGTG	TGGGGCTCCT	6600
GCTTCTGTTC	TTCCCGTGCA	GCCTTGGGAA	TGTTGGAGGC	CGTATCCAGG	GACCCCTGGG	6660
CCTCCTGGGA	TGGCCTCTCT	GGATCAGCCT	TGGAAGGTTC	CAGGCTGCCC	TTAGGCTCCC	6720
ACATTCTTCC	CCAGTCACGC	TCTCCTCGCC	CTGCCCACAC	CAGTCCTGTG	ACCTTGCCT	6780
GAGTTGTGAC	TTCCCACCCC	TCCCCGGCCT	AGAGGAAAGC	TGCCTGGCCC	CTCAGTGGGA	6840
CTCCCGCCCA	CTGACCCTCT	GTCCACCATA	CACAGACAGG	GGCACTATCC	ACAAGGTGGT	6900
GGAACCGGGG	GAGCAGGAGC	ACAGCTTCGC	CTTCAACATC	ATGGAGATCC	AGCCCTTCCG	6960

CCGCGCGGCT	GCCATCCAGA	CCATGTCGCT	GGATGCTGAG	CGGGTGAGCC	TTCCCCCACT	7020
GCGTCCCATG	GGCTATGCAG	TGACTGCAGC	TGAGGACAGG	GCTCCTTTGC	ATGTGATTG	7080
TGTGTTCTTT	TAAGAGCTTC	TAGGCCTTAG	GGCCTGGACA	TTTAGGACTG	AGTGTGGGGT	7140
GGGGCCCGGG	CCTGACCCAA	TCCTGCTGTC	CTTCCAGAGG	AAGCTGTATG	TGAGCTCCCA	7200
GTGGGAGGTG	AGCCAGGTGC	CCCTGGACCT	GTGTGAGGTC	TATGGCGGGG	GCTGCCACGG	7260
TTGCCTCATG	TCCCGAGACC	CCTACTGCGG	CTGGGACCAG	GGCCGCTGCA	TCTCCATCTA	7320
CAGCTCCGAA	CGGTACGTTG	GCCGGGATCC	CTCCGTCCCT	GGGACAAGGT	GGGCATGGGA	7380
CAGGGGGAGG	TGTTGTCGGG	CTGGAAGAGG	TGGCGGTACT	GGGCCTTTCT	TGTGGGACCT	7440
CCTCTCTACT	GGAAGTGCAC	TAGGGGTAAG	GATATGAGGG	TCAGGTCTGC	AGCCTTGTAT	7500
CTGCTGATCC	TCTTTGTC	TTCCCACTCC	AGGTCAGTGC	TGCAATCCAT	TAATCCAGCC	7560
GAGCCACACA	AGGAGTGTCC	CAACCCAAA	CCAGGTACCT	GATCTGGCCC	TGCTGGCGGC	7620
TGTGGCCCAA	TGAGTGGGGT	ACTGCCCTGC	CCTGATTGTC	CTGGTCTGAG	GGAAACATGG	7680
CCTTGTCCTG	TGGGCCCCAG	GTACATGGGG	CAGGATACAG	TCCTGCAGAG	GGAGCCCTCT	7740
TGGTGGGATG	AGCGAGACGG	GAGAAAAAAG	GAGGACGCTG	AGGGCTGGGT	TCCCCACGTT	7800
CATTGAGAAG	CCTTGTCCTG	GGATCCCAGT	CGGTGGGGAG	GACACATCCT	CCCCTGGGAG	7860
CTCTTTGTCC	CTCCTCACGG	CTGCTTCCCC	ACTGCCTCCC	CAGACAAGGC	CCCACTGCAG	7920
AAGGTTTCCC	TGGCCCCAAA	CTCTCGCTAC	TACCTGAGCT	GCCCCATGGA	ATCCCGCCAC	7980
GCCACCTACT	CATGGCGCCA	CAAGGAGAAC	GTGGAGCAGA	GCTGCGAACC	TGGTCACCAG	8040
AGCCCCAACT	GCATCCTGTT	CATCGAGAAC	CTCACGGCGC	AGCAGTACGG	CCACTACTTC	8100
TGCGAGGCCC	AGGAGGGCTC	CTACTTCCGC	GAGGCTCAGC	ACTGGCAGCT	GCTGCCCCGAG	8160
GACGGCATCA	TGGCCGAGCA	CCTGCTGGGT	CATGCCTGTG	CCCTGGCCGC	CTCCCTCTGG	8220
CTGGGGGTGC	TGCCCACACT	CACTCTTGGC	TTGCTGGTCC	ACTAGGGCCT	CCCGAGGCTG	8280
GGCATGCCTC	AGGCTTCTGC	AGCCAGGGC	ACTAGAACGT	CTCACACTCA	GAGCCGGCTG	8340
GCCCCGGAGC	TCCTTGCTG	CCACTTCTTC	CAGGGGACAG	AATAACCCAG	TGGAGGATGC	8400
CAGGCCTGGA	GACGTCCAGC	CGCAGGCGGC	TGCTGGGCCC	CAGGTGGCGC	ACGGATGGTG	8460
AGGGGCTGAG	AATGAGGGCA	CCGACTGTGA	AGCTGGGGCA	TCGATGACCC	AAGACTTTAT	8520
CTTCTGGAAA	ATATTTTCA	GACTCCTCAA	ACTTGACTAA	ATGCAGCGAT	GCTCCCAGCC	8580
CAAGAGCCCA	TGGGTCGGGG	AGTGGGTTTG	GATAGGAGAG	CTGGGACTCC	ATCTCGACCC	8640
TGGGGCTGAG	GCCTGAGTCC	TTCTGGACTC	TTGGTACCCA	CATTGCCTCC	TTCCCCTCCC	8700

TCTCTCATGG CTGGGTGGCT GGTGTTCTTG AAGACCCAGG GCTACCCTCT GTCCAGCCCT 8760
 GTCCTCTGCA GCTCCCTCTC TGGTCCTGGG TCCCACAGGA CAGCCGCCTT GCATGTTTAT 8820
 TGAAGGATGT TTGCTTTCCG GACGGAAGGA CGGAAAAAGC TCTGAAAAAA AAAAAAAAAA 8880
 AAAAAAAAAA 8888

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6622 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GATATCATGG AGATAATTAA AATGATAACC ATCTCGCAA TAAATAAGTA TTTTACTGTT 60
 TTCGTAACAG TTTTGTAATA AAAAAACCTA TAAATATGAA ATTCTTAGTC AACGTTGCCC 120
 TTGTTTTTAT GGTCGTATAC ATTTCTTACA TCTATGCGGA TCGATGGGGA TCCGCCCAGG 180
 GCCACCTAAG GAGCGGACCC CGCATCTTCG CCGTCTGGAA AGGCCATGTA GGGCAGGACC 240
 GGGTGGACTT TGGCCAGACT GAGCCGCACA CCGTGCTTTT CCACGAGCCA GGCAGCTCCT 300
 CTGTGTGGGT GGGAGGACGT GGCAAGGTCT ACCTCTTTGA CTTCCCCGAG GGCAAGAACG 360
 CATCTGTGCG CACGGTGAAT ATCGGCTCCA CAAAGGGGTC CTGTCTGGAT AAGCGGGACT 420
 GCGAGAACTA CATCACTCTC CTGGAGAGGC GGAGTGAGGG GCTGCTGGCC TGTGGCACCA 480
 ACGCCCGGCA CCCCAGCTGC TGGAACCTGG TGAATGGCAC TGTGGTGCCA CTTGGCGAGA 540
 TGAGAGGCTA TGCCCCCTTC AGCCCGGACG AGAACTCCCT GGTTCTGTTT GAAGGGGACG 600
 AGGTGTATTC CACCATCCGG AAGCAGGAAT ACAATGGGAA GATCCCTCGG TTCCGCCGCA 660
 TCCGGGGCGA GAGTGAGCTG TACACAGTG ATACTGTCAT GCAGAACCCA CAGTTCATCA 720
 AAGCCACCAT CGTGCACCAA GACCAGGCTT ACGATGACAA GATCTACTAC TTCTTCCGAG 780
 AGGACAATCC TGACAAGAAT CCTGAGGCTC CTCTCAATGT GTCCCGTGTG GCCCAGTTGT 840
 GCAGGGGGGA CCAGGGTGGG GAAAGTTCAC TGTCAGTCTC CAAGTGGAAC ACTTTTCTGA 900
 AAGCCATGCT GGTATGCAGT GATGCTGCCA CCAACAAGAA CTTCAACAGG CTGCAAGACG 960
 TCTTCCTGCT CCCTGACCCC AGCGGCCAGT GGAGGGACAC CAGGGTCTAT GGTGTTTTCT 1020

CCAACCCCTG	GAAC TACTCA	GCCGTCTGTG	TGTATTCCCT	CGGTGACATT	GACAAGGTCT	1080
TCCGTACCTC	CTCACTCAAG	GGCTACCACT	CAAGCCTTCC	CAACCCGCGG	CCTGGCAAGT	1140
GCCTCCCAGA	CCAGCAGCCG	ATACCCACAG	AGACCTTCCA	GGTGGCTGAC	CGTCACCCAG	1200
AGGTGGCGCA	GAGGGTGGAG	CCCATGGGGC	CTCTGAAGAC	GCCATTGTTC	CACTCTAAAT	1260
ACCACTACCA	GAAAGTGGCC	GTTCAACGCA	TGCAAGCCAG	CCACGGGGAG	ACCTTTCATG	1320
TGCTTTACCT	AACTACAGAC	AGGGGCACTA	TCCACAAGGT	GGTGGAACCG	GGGGAGCAGG	1380
AGCACAGCTT	CGCCTTCAAC	ATCATGGAGA	TCCAGCCCTT	CCGCCGCGCG	GCTGCCATCC	1440
AGACCATGTC	GCTGGATGCT	GAGCGGAGGA	AGCTGTATGT	GAGCTCCCAG	TGGGAGGTGA	1500
GCCAGGTGCC	CCTGGACCTG	TGTGAGGTCT	ATGGCGGGGG	CTGCCACGGT	TGCCTCATGT	1560
CCCGAGACCC	CTACTGCGGC	TGGGACCAGG	GCCGCTGCAT	CTCCATCTAC	AGCTCCGAAC	1620
GGTCAGTGCT	GCAATCCATT	AATCCAGCCG	AGCCACACAA	GGAGTGTCCT	AACCCCAAAC	1680
CAGACAAGGC	CCCACTGCAG	AAGGTTTCCC	TGGCCCCAAA	CTCTCGCTAC	TACCTGAGCT	1740
GCCCCATGGA	ATCCCGCCAC	GCCACCTACT	CATGGCGCCA	CAAGGAGAAC	GTGGAGCAGA	1800
GCTGCGAACC	TGGTCACCAG	AGCCCCAACT	GCATCCTGTT	CATCGAGAAC	CTCACGGCGC	1860
AGCAGTACGG	CCACTACTTC	TGCGAGGCCC	AGGAGGGCTC	CTACTTCCGC	GAGGCTCAGC	1920
ACTGGCAGCT	GCTGCCCCGAG	GACGGCATCA	TGGCCGAGCA	CCTGCTGGGT	CATGCCTGTG	1980
CCCTGGCTGC	CTGAATTCGA	AGCTTGGAGT	CGACTCTGCT	GAAGAGGAGG	AAATTCTCCT	2040
TGAAGTTTCC	CTGGTGTTCA	AAGTAAAGGA	GTTTGACCA	GACGCACCTC	TGTTCACTGG	2100
TCCGGCGTAT	TAAAACACGA	TACATTGTTA	TTAGTACATT	TATTAAGCGC	TAGATTCTGT	2160
GCGTTGTTGA	TTTACAGACA	ATTGTTGTAC	GTATTTTAAT	AATTCATTAA	ATTTATAATC	2220
TTTAGGGTGG	TATGTTAGAG	CGAAAATCAA	ATGATTTTCA	GCGTCTTTAT	ATCTGAATTT	2280
AAATATTAAA	TCCTCAATAG	ATTTGTAAAA	TAGGTTTCGA	TTAGTTTCAA	ACAAGGGTTG	2340
TTTTTCCGAA	CCGATGGCTG	GACTATCTAA	TGGATTTTCG	CTCAACGCCA	CAAACTTGC	2400
CAAATCTTGT	AGCAGCAATC	TAGCTTTGTC	GATATTCGTT	TGTGTTTTGT	TTTGTAATAA	2460
AGGTTTCGACG	TCGTTCAAAA	TATTATGCGC	TTTTGTATTT	CTTTCATCAC	TGTCGTTAGT	2520
GTACAATTGA	CTCGACGTAA	ACACGTTAAA	TAAAGCCTGG	ACATATTTAA	CATCGGGCGT	2580
GTTAGCTTTA	TTAGGCCGAT	TATCGTCGTC	GTCCCAACCC	TCGTCGTTAG	AAGTTGCTTC	2640
CGAAGACGAT	TTTGCCATAG	CCACACGACG	CCTATTAATT	GTGTCGGCTA	ACACGTCCGC	2700

GATCAAATTT	GTAGTTGAGC	TTTTTGAAT	TATTTCTGAT	TGCGGGCGTT	TTTGGGCGGG	2760
TTTCAATCTA	ACTGTGCCCC	ATTTTAATTC	AGACAACACG	TTAGAAAGCG	ATGGTGCAGG	2820
CGGTGGTAAC	ATTTCAAGACG	GCAAATCTAC	TAATGGCGGC	GGTGGTGGAG	CTGATGATAA	2880
ATCTACCATC	GGTGGAGGCG	CAGGCGGGGC	TGGCGGCGGA	GGCGGAGGCG	GAGGTGGTGG	2940
CGGTGATGCA	GACGGCGGTT	TAGGCTCAAA	TTGTCTCTTT	CAGGCAACAC	AGTCGGCACC	3000
TCAACTATTG	TACTGGTTTC	GGGCGTATGG	TGCACTCTCA	GTACAATCTG	CTCTGATGCC	3060
GCATAGTTAA	GCCAGCCCCG	ACACCCGCCA	ACACCCGCTG	ACGCGCCCTG	ACGGGCTTGT	3120
CTGCTCCCGG	CATCCGCTTA	CAGACAAGCT	GTGACCGTCT	CCGGGAGCTG	CATGTGTCAG	3180
AGGTTTTTAC	CGTCATCACC	GAAACGCGCG	AGACGAAAGG	GCCTCGTGAT	ACGCCTATTT	3240
TTATAGGTTA	ATGTCATGAT	AATAATGGTT	TCTTAGACGT	CAGGTGGCAC	TTTTCGGGGA	3300
AATGTGCGCG	GAACCCCTAT	TTGTTTATTT	TTCTAAATAC	ATTCAAATAT	GTATCCGCTC	3360
ATGAGACAAT	AACCCTGATA	AATGCTTCAA	TAATATTGAA	AAAGGAAGAG	TATGAGTATT	3420
CAACATTTCC	GTGTCGCCCT	TATTCCCTTT	TTTGCGGCAT	TTTGCCCTCC	TGTTTTTGCT	3480
CACCCAGAAA	CGCTGGTGAA	AGTAAAAGAT	GCTGAAGATC	AGTTGGGTGC	ACGAGTGGGT	3540
TACATCGAAC	TGGATCTCAA	CAGCGGTAAAG	ATCCTTGAGA	GTTTTCGCCC	CGAAGAACGT	3600
TTTCCAATGA	TGAGCACTTT	TAAAGTTCTG	CTATGTGGCG	CGGTATTATC	CCGTATTGAC	3660
GCCGGGCAAG	AGCAACTCGG	TCGCCGCATA	CACTATTCTC	AGAATGACTT	GGTTGAGTAC	3720
TCACCAGTCA	CAGAAAAGCA	TCTTACGGAT	GGCATGACAG	TAAGAGAATT	ATGCAGTGCT	3780
GCCATAACCA	TGAGTGATAA	CACTGCGGCC	AACTTACTTC	TGACAACGAT	CGGAGGACCG	3840
AAGGAGCTAA	CCGCTTTTTT	GCACAACATG	GGGGATCATG	TAACTCGCCT	TGATCGTTGG	3900
GAACCGGAGC	TGAATGAAGC	CATACCAAAC	GACGAGCGTG	ACACCACGAT	GCCTGTAGCA	3960
ATGGCAACAA	CGTTGCGCAA	ACTATTAAGT	GGCGAACTAC	TTACTCTAGC	TTCCCGGCAA	4020
CAATTAATAG	ACTGGATGGA	GGCGGATAAA	GTTGCAGGAC	CACTTCTGCG	CTCGGCCCTT	4080
CCGGCTGGCT	GGTTTATTGC	TGATAAATCT	GGAGCCGGTG	AGCGTGGGTC	TCGCGGTATC	4140
ATTGCAGCAC	TGGGGCCAGA	TGGTAAGCCC	TCCCGTATCG	TAGTTATCTA	CACGACGGGG	4200
AGTCAGGCAA	CTATGGATGA	ACGAAATAGA	CAGATCGCTG	AGATAGGTGC	CTCACTGATT	4260
AAGCATTGGT	AACTGTCAGA	CCAAGTTTAC	TCATATATAC	TTTAGATTGA	TTTAAAACTT	4320
CATTTTTAAT	TTAAAAGGAT	CTAGGTGAAG	ATCCTTTTTG	ATAATCTCAT	GACCAAAATC	4380
CCTTAACGTG	AGTTTTTCGT	CCACTGAGCG	TCAGACCCCG	TAGAAAAGAT	CAAAGGATCT	4440

TCTTGAGATC CTTTTTTTCT GCGCGTAATC TGCTGCTTGC AAACAAAAAA ACCACCGCTA	4500
CCAGCGGTGG TTTGTTTGCC GGATCAAGAG CTACCAACTC TTTTCCGAA GGTAACGGC	4560
TTCAGCAGAG CGCAGATACC AAATACTGTT CTTCTAGTGT AGCCGTAGTT AGGCCACCAC	4620
TTCAAGAACT CTGTAGCACC GCCTACATAC CTCGCTCTGC TAATCCTGTT ACCAGTGGCT	4680
GCTGCCAGTG GCGATAAGTC GTGTCTTACC GGGTTGGACT CAAGACGATA GTTACCGGAT	4740
AAGGCGCAGC GGTCGGGCTG AACGGGGGGT TCGTGACAC AGCCAGCTT GGAGCGAACG	4800
ACCTACACCG AACTGAGATA CCTACAGCGT GAGCTATGAG AAAGCGCCAC GCTTCCCGAA	4860
GGGAGAAAGG CGGACAGGTA TCCGGTAAGC GGCAGGGTCG GAACAGGAGA GCGCACGAGG	4920
GAGCTTCCAG GGGGAAACGC CTGGTATCTT TATAGTCCTG TCGGGTTTCG CCACCTCTGA	4980
CTTGAGCGTC GATTTTTGTG ATGCTCGTCA GGGGGGCGGA GCCTATGGAA AAACGCCAGC	5040
AACGCGCCT TTTTACGGTT CCTGGCCTTT TGCTGGCCTT TTGCTCACAT GTTCTTTCCT	5100
GCGTTATCCC CTGATTCTGT GGATAACCGT ATTACCGCCT TTGAGTGAGC TGATACCGCT	5160
CGCCGCAGCC GAACGACCGA GCGCAGCGAG TCAGTGAGCG AGGAAGCATC CTGCACCATC	5220
GTCTGCTCAT CCATGACCTG ACCATGCAGA GGATGATGCT CGTGACGGTT AACGCCTCGA	5280
ATCAGCAACG GCTTGCCGTT CAGCAGCAGC AGACCATTTT CAATCCGCAC CTCGCGGAAA	5340
CCGACATCGC AGGCTTCTGC TTCAATCAGC GTGCCGTGCG CGGTGTGCAG TTCAACCACC	5400
GCACGATAGA GATTCCGGGAT TTCGGCGCTC CACAGTTTCG GGTTTTCGAC GTTCAGACGT	5460
AGTGTGACGC GATCGGTATA ACCACCACGC TCATCGATAA TTTACCCGCC GAAAGGCGCG	5520
GTGCCGCTGG CGACCTGCGT TTCACCCTGC CATAAAGAAA CTGTTACCCG TAGGTAGTCA	5580
CGCAACTCGC CGCACATCTG AACTTCAGCC TCCAGTACAG CGCGGCTGAA ATCATCATT	5640
AAGCGAGTGG CAACATGGAA ATCGCTGATT TGTGTAGTCG GTTTATGCAG CAACGAGACG	5700
TCACGGAAAA TGCCGCTCAT CCGCCACATA TCCTGATCTT CCAGATAACT GCCGTCACTC	5760
CAACGCAGCA CCATCACCGC GAGGCGGTTT TCTCCGGCGC GTAAAAATGC GCTCAGGTCA	5820
AATTGACAGC GCAAACGACT GTCCTGGCCG TAACCGACCC AGCGCCCGTT GCACCACAGA	5880
TGAAACGCCG AGTTAACGCC ATCAAAAATA ATTGCGTCTT GGCCTTCCTG TAGCCAGCTT	5940
TCATCAACAT TAAATGTGAG CGAGTAACAA CCCGTCGGAT TCTCCGTGGG AACAAACGGC	6000
GGATTGACCG TAATGGGATA GGTCACGTTG GTGTAGATGG GCGCATCGTA ACCGTGCATC	6060
TGCCAGTTTG AGGGGACGAC GACAGTATCG GCCTCAGGAA GATCGCACTC CAGCCAGCTT	6120

TCCGGCACCG CTTCTGGTGC CGGAAACCAG GCAAAGCGCC ATTCGCCATT CAGGCTGCGC 6180
 AACTGTTGGG AAGGGCGATC GGTGCGGGCC TCTTCGCTAT TACGCCAGCT GGCGAAAGGG 6240
 GGATGTGCTG CAAGGCGATT AAGTTGGGTA ACGCCAGGGT TTTCCCAGTC ACGACGTTGT 6300
 AAAACGACGG GATCTATCAT TTTTAGCAGT GATTCTAATT GCAGCTGCTC TTTGATACAA 6360
 CTAATTTTAC GACGACGATG CGAGCTTTTA TTCAACCGAG CGTGCATGTT TGCAATCGTG 6420
 CAAGCGTTAT CAATTTTTC AATTATCGTATT GTTGACATC AACAGGCTGG ACACCACGTT 6480
 GAACTCGCCG CAGTTTTGCG GCAAGTTGGA CCCGCCGCGC ATCCAATGCA AACTTTCCGA 6540
 CATTCTGTTG CCTACGAACG ATTGATTCTT TGTCCATTGA TCGAAGCGAG TGCCTTCGAC 6600
 TTTTTCGTGT CCAGTGTGGC TT 6622

(2) INFORMATION FOR SEQ ID NO:43:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

CCGGATCCGC CCAGGGCCAC CTAAGGAGCG G

31

(2) INFORMATION FOR SEQ ID NO:44:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 29 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

CTGAATTCAG GAGCCAGGGC ACAGGCATG

29